ELECTRON TRANSPORT IN LOW DIMENSIONAL DISORDERED SYSTEMS

by

Michael John Kearney

A thesis

presented to the University of Warwick

in partial fulfilment of the requirements

for entry to the degree of

Doctor of Philosophy

Department of Physics

May 1988

Abstract

The transport properties of low dimensional systems (especially wires) are investigated when the dominant scattering is due to the impurities and is elastic. Such a situation is expected to be relevant to experiments carried out at very low (liquid Helium) temperatures.

Initially a Boltzmann formalism is used to illustrate the effects of multiple sub-band occupancy. Structure is found in the electrical conductivity, thermal conductivity and thermopower when plotted as a function of chemical potential, due to the lateral quantisation of the electron states. These quantum size effects (QSE) are most pronounced in the thermopower, which is expected to show sign changes when the chemical potential sweeps through a sub-band minimum.

A more sophisticated treatment based on Green's function methods reveals the importance of lifetime broadening in quasi-one-dimensional systems, which smears out the single-particle density of states and the QSE. The type of behaviour expected of realistic devices is explored, and it is shown that the thermopower offers the best chance of observing confinement effects.

The formal theory may also be applied to the weak localisation corrections in multi-sub-band systems. An expression for the correction term is obtained which is valid for arbitrary channel width, and enables the crossover from a linear to logarithmic scaling in L_ϕ to be demonstrated. A transverse inelastic length is derived, and shown to be the length scale which controls the system dimensionality rather than L_ϕ. The implication for experiment in narrow channels is discussed.

Weak localisation corrections are also calculated for the thermopower and the thermal conductivity. This corrects a result due to Ting et al (1982) that there are no weak localisation corrections to the thermopower in 2D. These results are shown to be a consequence of a rather general scaling theory of thermal transport which has wider implications, such as for the behaviour expected near a Metal-Insulator transition for example. Comparison with the single parameter scaling theory of the zero temperature conductance is made. Fluctuation effects for thermal transport in mesoscopic samples are also explored (both numerically and analytically), and the analogue of universal conductance fluctuations explicitly demonstrated.

Table of Contents

LIST OF ILLUSTRATIONS

Acknowledgements

I would like to take this opportunity to thank some of the many people who have helped me reach this stage.

My fiancée and parents for their support and encouragement

Professor P.N.Butcher for his excellent supervision

My colleagues at Warwick, GEC and to all the people who have shown friendship and taken a helpful interest in this work

The SERC and the GEC Hirst research centre for financial support through a CASE award.

Finally, to all those who enabled me to attend conferences both in this country and abroad.

DECLARATION

This thesis contains an account of my own independent research work performed in the Department of Physics at the University of Warwick between October 1985 and May 1988 under the supervision of Professor Paul N Butcher.

Some of the work has previously been published as follows:-

1. "A calculation of the effect of sub-band structure on the thermopower of a quasi-one-dimensional wire"

M J Kearney and P N Butcher 1986 J. Phys. C 19 5429

2. "The effect of lifetime broadening on the conductivity and thermopower of a quasi-1D wire"

M J Kearney and P N Butcher 1987 J. Phys. C 20 47

3. "Thermal transport in disordered systems"

M J Kearney and P N Butcher 1988 J. Phys. C 21 L265

4. "Weak localisation in the quasi-one-dimensional regime with many occupied sub-bands"

M J Kearney and P N Butcher 1988 accepted for publication and to appear in J. Phys. C

To Sally

CHAPTER 1

SIMPLE CONCEPTS OF LOW DIMENSIONAL PHYSICS.

1.1 Introduction.

This thesis represents an attempt to study theoretically some of the features of electron transport expected in low dimensional systems. The easiest way to envisage how low dimensionality may arise is to imagine a cube of metal or doped semiconductor, and consider the role the boundaries play in determining its behaviour as a function of the side length L. When L is large, the customary view is that the boundaries are sufficiently far away to have no effect on the local behaviour. This is commonly referred to as the bulk, or three dimensional limit, for obvious reasons. As L becomes smaller however, there will obviously come a point when the presence of the boundaries can no longer be ignored. If L is of the order of the mean free path of the conduction electrons for example, then boundary scattering may be important which in turn will affect the resistance of the solid. Decreasing L still further will have progressively more effect until when L is of the order of the de Broglie wavelength λ of the electrons they can no longer be regarded as free particles. Momentum is not a good quantum number in such a situation and not surprisingly, the physical properties become drastically different from their bulk behaviour.

Instead of a cubic sample, we can now imagine thin sheets or wires where the confinement occurs in only one or two directions respectively. If this confinement is on the scale of λ, then the number of dimensions in which the conduction electrons are essentially free to move is only two or one accordingly. This then defines the effective dimensionality of the system from the point of view of electronic transport. These low dimensional systems should have interesting properties which are quite different to those encountered in bulk solids on account of the confinement of the electrons. Studying transport in such systems from a theoretical point of view is therefore expected to be both exciting and challenging. We shall find that the behaviour is not just quantitatively different from that of the bulk, but also highlights new phenomena unique to these small systems, especially in wires where the confinement effects are greatest.

1.2 Transport in bulk solids.

Before commencing with a study of low dimensional transport, it is helpful to recall some corresponding results for three dimensional solids. This is useful both in providing similarities and underlining differences when and where they arise. We have in mind metals and doped semiconductors which have a partially filled conduction band, so that the dominant contribution to the transport comes from conduction electrons which are essentially free to move throughout the solid. Three particularly useful quantities to consider (both experimentally and theoretically) are the electrical conductivity, the thermal conductivity and the thermoelectricpower, or thermopower. Many aspects of the theory of these coefficients when conduction electrons dominate the transport process are now rather well understood in bulk materials (Butcher 1986, Blatt 1968, Nag 1980, Smith et al 1967). For the purpose of illustration we shall consider a simple model of transport which is pertinent to the low temperature scattering (elastic) of free particles by randomly placed impurities. This so-called Sommerfeld-Boltzmann theory is described in many sources (see e.g. Ashcroft and Mermin 1981).

Perhaps the easiest transport phenomenon to understand (at least conceptually) is that of electrical conduction, where a current flows in a solid in response to an applied electric field. The current is carried by the conduction electrons (since they are free to move), and remains finite because scattering tends to degrade it. The constant of proportionality relating the current to the field, assuming the applied field is small, is the electrical conductivity σ. In the simple Sommerfeld model it has the form,

$$\sigma = \frac{ne^2\tau}{m^*} \tag{1.1}$$

where n is the (conduction) electron density, m^* is the conduction band effective mass and τ is a timescale characteristic of the scattering. This simple expression describes many different systems reasonably well and more sophisticated formulae obviously give better results. The salient features of this expression are that σ is finite at $T = 0$, and through depending linearly upon τ, is inversely proportional to the rate at which the electrons scatter. This conductivity is a property of a given sample, and although it depends upon the actual impurity concentration, it does not

depend upon the size or the shape of the sample. One of the first things we shall find when investigating low dimensional systems is that this is no longer the case.

Conduction electrons can also carry heat as well as electrical current. The thermal conductivity κ is a measure of the rate at which this heat can be transported through a solid. A similar level of calculation to the above gives for the electronic contribution to κ at low temperatures (Ashcroft and Mermin 1981),

$$\kappa = \frac{\pi^2}{3}\frac{k_B^2 T n \tau}{m^*} \tag{1.2}$$

At zero temperature there can be no heat transport and so $\kappa(T=0)$ vanishes. The interesting point about equations (1.1) and (1.2) is that $\kappa/\sigma T$ is apparently a universal number (the Lorenz number); a relationship known as the Wiedemann-Franz law,

$$\frac{\kappa}{\sigma T} = \frac{\pi^2}{3}\frac{k_B^2}{e^2} = 2.44 \times 10^{-8}\ Watt\text{-}Ohm/K^2 \tag{1.3}$$

which although only derived crudely, describes a number of solids surprisingly well (Ashcroft and Mermin 1981). Deviations from the Wiedemann-Franz law at higher temperatures due to inelastic scattering can also be accounted for. It is important to stress that the second law of thermodynamics demands that κ is always a positive quantity.

The third commonly measured transport coefficient is the thermopower S. When a temperature gradient is applied across an isolated sample where no current can flow, a potential difference will in general appear across its ends. The thermopower is the constant of proportionality relating the resulting electric field to this temperature gradient. At low temperatures, the Mott (Cutler and Mott 1969) formula for an electron gas is,

$$S = -\frac{\pi^2}{3}\frac{k_B^2 T}{|e|\varepsilon_F} \tag{1.4}$$

which, intriguingly, is independent of the relaxation time τ. The Mott formula predicts that as $T \to 0$, S is negative and should vanish linearly with T. Experimentally however this expression is often found to be drastically wrong, sometimes even having the incorrect sign. The thermopower is possibly the least well understood of all transport coefficients, being notoriously sensitive to microscopic details of the system (Mahan 1981) and being complicated by the presence of

phonon-drag (Blatt 1968), which relates to the non-equilibrium nature of the phonons in the presence of a temperature gradient. Excellent accounts of the problems involved in treating thermopower theoretically have been given by Herring (1954), Bailyn (1967), Vilenkin and Taylor (1978), Sivan and Imry (1986) and Stedman and Kaiser (1987). The sensitivity to detail, although complicating the theory, is perhaps the most important reason why thermopower is studied, because of the information it can potentially yield about the way solids behave. We shall see that this is equally applicable to low dimensional systems as well, where thermal transport has received comparatively little attention.

1.3 The Silicon MOSFET.

Easily the most versatile example of a two dimensional electron system is the MOSFET (Metal Oxide Semiconductor Field Effect Transistor), so it is convenient to begin a discussion with a look at this system. A schematic drawing of an n-channel silicon device is shown in figure (1.1). Current can only flow between the source and the drain if a positive voltage is applied to the metal gate. If this voltage is sufficiently large, the resulting band-bending leads to a confining potential which binds the electrons to a thin layer near the $Si-SiO_2$ interface (figure (1.2)). This n-type inversion layer is responsible for carrying the current in the device. The width of the inversion layer depends upon the gate voltage, but is typically very narrow, of the order of 3 - 5 nm. As a result, the electrons are free to move in two directions (parallel to the interface), but not in the third. Such a device, when conducting, exhibits many of the properties one would associate with a purely 2D system (for a comprehensive overview see the review by Ando et al 1982).

The first interesting question concerns the nature of the electron states in the device. Electrons moving in the inversion layer of a MOSFET obviously experience a complicated potential, however systematic studies (Ando et al 1982 and references therein) have revealed several important facts which simplify analysis of the problem considerably. Firstly, the periodic potential due to the Silicon atoms is rather well described within the confines of effective mass theory (Butcher 1973, Smith et al 1967), and so in the plane of the inversion layer, the conduction electrons can be treated as plane waves (at least to a first approximation). Transverse to the inversion layer, the

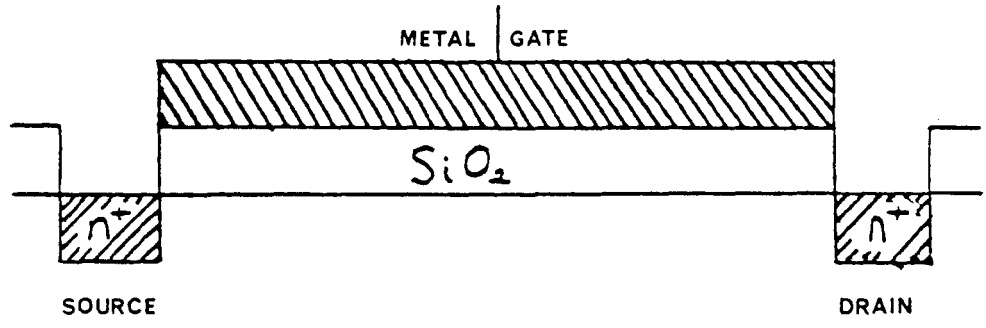


Figure 1.1: A schematic diagram of an n-channel silicon MOSFET.

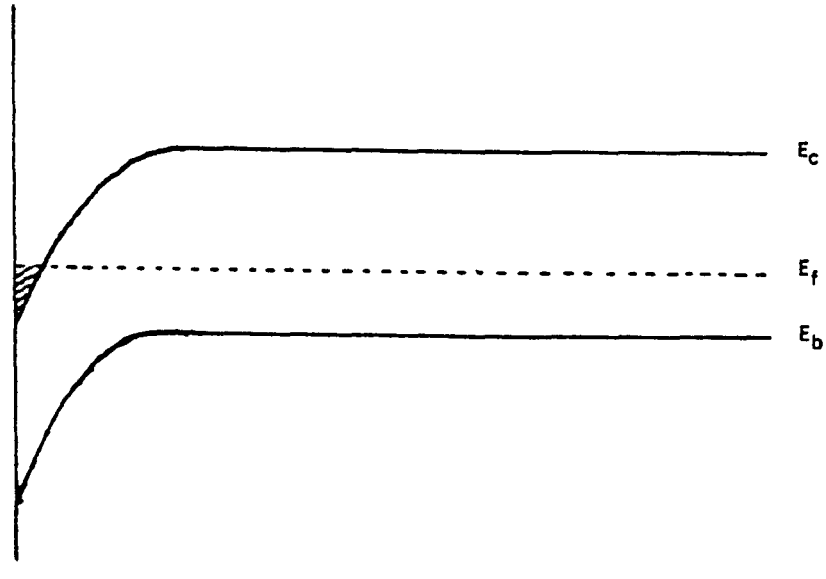


Figure 1.2: The gate voltage produces band bending until the conduction band drops below the Fermi level, leading to the creation of the inversion layer (shaded region). E_c and E_b are the conduction and valence band edges respectively.

electrons experience a confining potential due to the combined effect of the gate voltage, interface potentials and coulomb interactions (Hartree term) with the other electrons. To find this potential one can solve the Schrodinger and Poisson equations in a self-consistent fashion. The net result is a confining potential sufficiently narrow to lead to quantisation perpendicular to the plane of the interface. If we consider an area A of the two dimensional electron system, the eigenfunctions are then simply,

$$\phi_\alpha(\mathbf{k}) = \left[\frac{1}{A}\right]^{\frac{1}{2}} e^{i\mathbf{k}.\mathbf{r}} \xi_\alpha(z) \tag{1.5}$$

where,

$$-\frac{\hbar^2}{2m^*}\frac{d^2}{dz^2}\xi_\alpha(z) + V(z)\,\xi_\alpha(z) = \varepsilon_\alpha\,\xi_\alpha(z). \tag{1.6}$$

In general, the eigenvalue $\mathbf{k}$ may be taken to be continuous, reflecting the translational symmetry in the plane of the interface. The eigenvalue ε_α on the other hand is discrete if the potential $V(z)$ is narrow. As an illustration, consider the model "box" potential,

$$\begin{aligned} V(z) &= 0 \quad 0<z<t \\ &= \infty \quad \textit{otherwise} \end{aligned}$$

when ε_α can take the values $\hbar^2\alpha^2\pi^2/(2m^*t^2)$ (α is a positive integer). When the Fermi energy $\varepsilon_F \approx \varepsilon_1$ the effect of lateral quantisation will be important, which occurs when the thickness t is of the order of the de Broglie wavelength $\approx 1/k_F$. The index α is known as the sub-band index, and each ε_α defines the bottom of a sub-band upon which a continuum of states is built from the plane-wave component of the wavefunction. A representation of the band structure arising in such a situation is shown in figure (1.3).

The quantisation due to spatial confinement leads to a peculiar "stair-case" structure in the density of states,

$$g(\varepsilon) = \sum_\alpha \frac{Am^*}{\pi\hbar^2}\,\theta(\varepsilon - \varepsilon_\alpha) \tag{1.7}$$

where $\theta(x)$ is the unit step function, and is shown in figure (1.4). When only the lowest sub-band is occupied the density of states is constant, and is exactly what one would associate with a purely two-dimensional system. This so called quantum limit (which is usually found in MOSFETS) helps to explain why they may be thought of as two-dimensional. Calculations based upon a

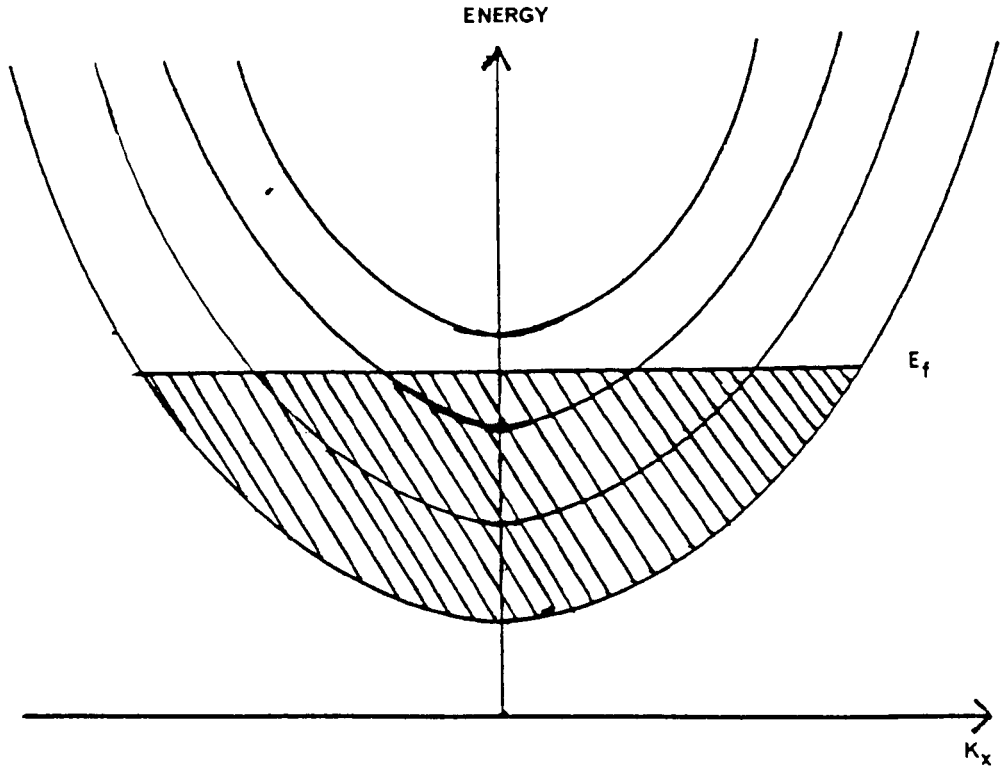


Figure 1.3: A representation of the sub-band structure for a given direction in k-space. The electrons occupy all the states up to the Fermi level (at $T = 0$).

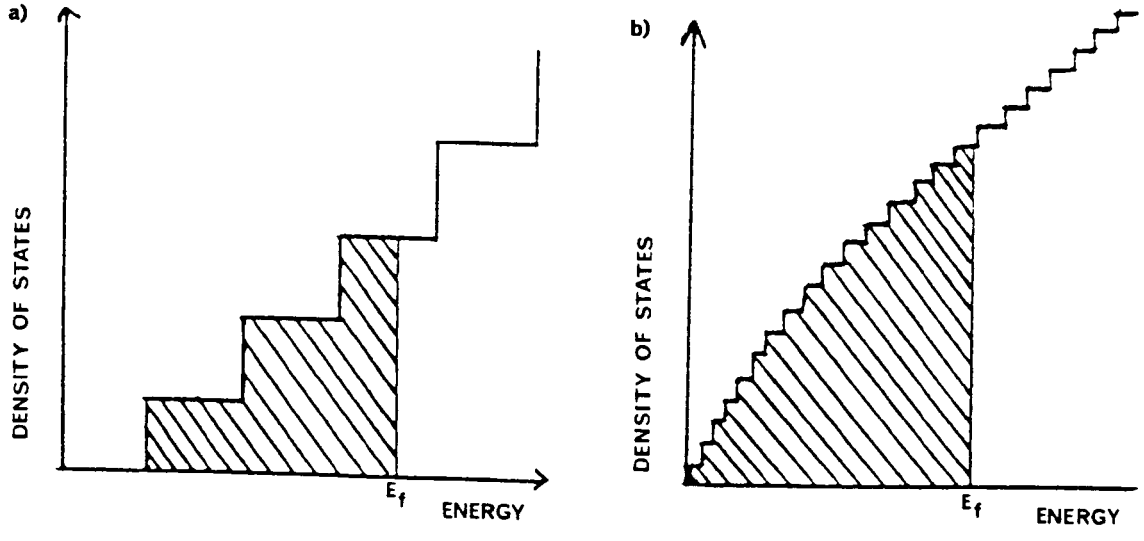


Figure 1.4: a) The density of states in a MOSFET with several occupied sub-bands. b) A wider MOSFET with many occupied sub-bands. The envelope of the steps varies as $\varepsilon^{1/2}$, characteristic of a 3D density of states.

model of a simple 2DEG have proved to be very successful in discussing MOSFET physics (Ando et al 1982). The most important feature of the system is that the electrons are essentially free to move in only two directions. Details of the exact form of the sub-band wavefunctions and the confining potential are, (at least from the point of view of transport properties) of lesser importance, which is helpful since it permits simple models to be used with relative accuracy. The same is also found to be true for the effects of exchange and correlation, and so the single-particle description is a reasonable basis with which to work. (There are of course situations where this is certainly not the case, and a many-body description must be invoked. Probably the most graphic illustration of this in a 2DEG is the fractional quantum hall effect (Laughlin 1983), which is inherently a many-body phenomenon).

By increasing the gate voltage it is possible to draw more electrons into the inversion layer and thus have more than one occupied sub-band. When several sub-bands are occupied we have a region of quasi-two-dimensionality, in which the electrons are still free to move in the plane of the layer, but inter-sub-band scattering generates an "effective" motion perpendicular to the interface. It is not hard to imagine that the singularities in the density of states (figure (1.4)) will lead to potentially novel behaviour in this regime, due for example to the changing scattering rates as a function of Fermi level. In later chapters we shall look at phenomena of this kind. The three dimensional limit is associated with a great many (closely spaced) sub-bands or equivalently, with confinement channels that are very wide. Just what constitutes "wide" is a pertinent question and will form a recurring theme throughout this work. The 2D to 3D crossover in the density of states is demonstrated graphically in figure (1.4). The great advantage of using MOSFETS to study 2D behaviour is that the Fermi level can be altered, which makes it possible to probe different regimes within the same sample. There are other examples of quasi-two-dimensional systems which have also proved useful in investigating 2D physics, and we mention some of these in passing. The GaAs - AlGaAs heterojunction can confine electrons in a small layer near the interface because of band bending in the presence of an abrupt compositional change. Details of the confining potential are different, but transport is similar to that found in MOSFETS (Ando et al 1982, Thornton et al 1986). Thin (deposited) films of metal and semiconductor often behave

two-dimensionally and have been especially useful in studying localisation phenomena (Bergmann 1984). Certain bulk solids can show aspects of low dimensionality due to strong spatial anisotropy in the system, the well defined planes in graphite being a good example. Finally, it has even proved possible to study electrons on the surface of liquid helium, where electrons have very high mobilities, and where crystallisation of a 2DEG has been observed at sufficiently low temperatures (Grimes and Adams 1979).

1.4 One dimensional systems.

Systems which are inherently one dimensional may be constructed in analogous fashion by spatially restricting the electrons in two dimensions instead of one. A sub-band structure will develop as in the MOSFET, although each sub-band will be labelled by two indices rather than one. The resulting density of states is even more peculiar,

$$g(\varepsilon) = \frac{L}{\pi\hbar}\left[\frac{m^*}{2}\right]^{\frac{1}{2}} \sum_{m,n} \frac{\theta(\varepsilon - \varepsilon_{mn})}{(\varepsilon - \varepsilon_{mn})^{\frac{1}{2}}} \qquad (1.8)$$

and is shown schematically in figure (1.5). The most interesting feature is the singularities which occur when the energy coincides with a sub-band minimum. As it stands, we can expect more exotic behaviour to arise in quasi- one-dimensional systems (compared to MOSFETS) because of these singularities. Devices to potentially probe such behaviour have been made in a variety of ways. Most of the early work concentrated upon constricting the 2DEG of a MOSFET in one more direction, either electrostatically (Fowler et al 1982), or by defining a narrow channel using electron beam (Skocpol et al 1982) or optical (Wheeler et al 1982) lithography. In a similar fashion, narrow channels may be defined in the 2DEG of a GaAs-AlGaAs heterojunction (Thornton et al 1986, van Houten et al 1988). As lithographic techniques have become more advanced, it has proved possible to fabricate wires of sufficiently small cross-sectional area to observe quasi-1D behaviour. A variety of materials have been used for this, such as AuPd alloy (Masden and Giordano 1982), aluminium (Santhanam et al 1984), lithium (Licini et al 1985) and n^+GaAs (Whittington et al 1986). Wires which are free standing in space have also been fabricated, from amorphous silicon nitride (Lee et al 1984) and AuPd (Smith et al 1985). Each of these devices is slightly different and highlights different aspects of one-dimensional transport. Typical

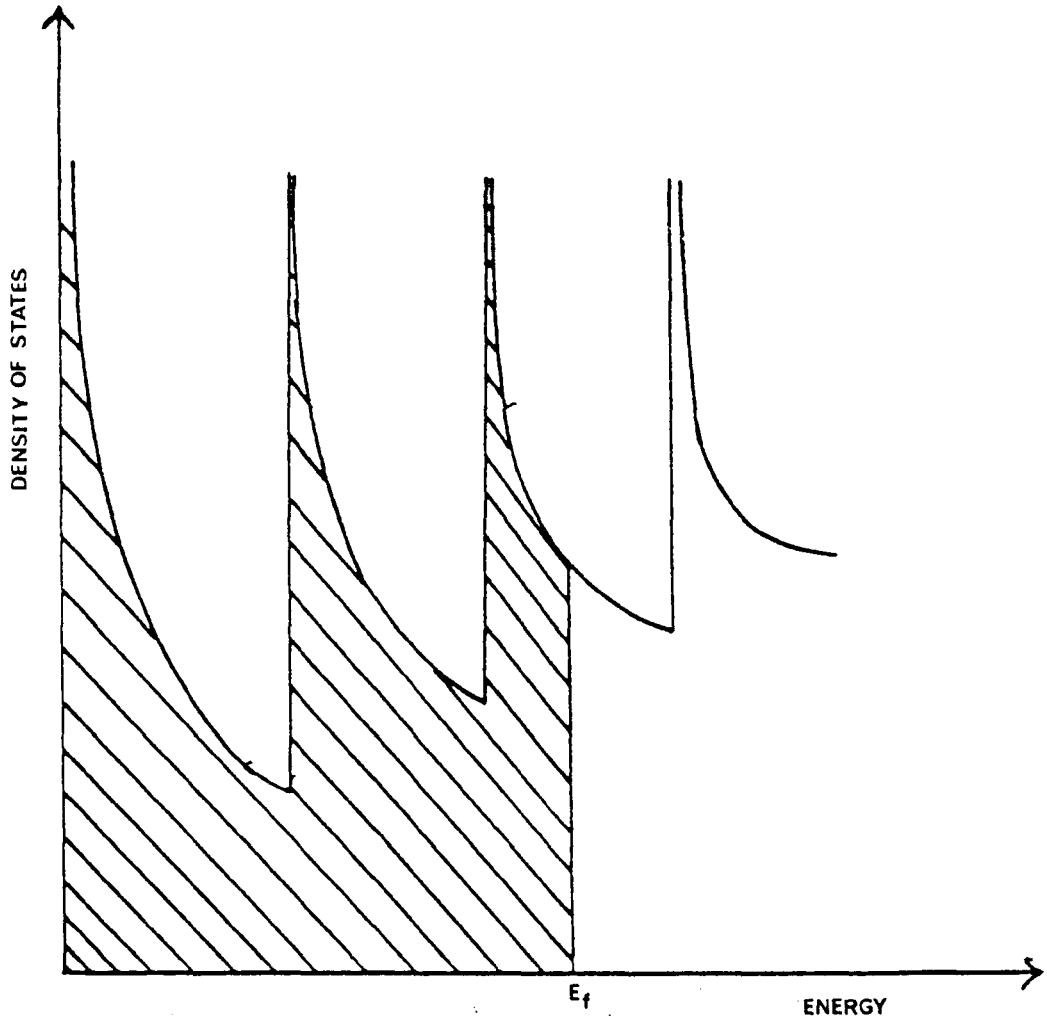


Figure 1.5: Quasi-one-dimensional density of states. The singularities vary as $(\varepsilon - \varepsilon_{mn})^{-1/2}$ and are integrable, so the electron density is finite.

dimensions are lengths measured in microns and lateral dimensions measured in some hundreds of angstroms; more will be said about this later. What is evident from the experiments is that these systems can behave very differently from the way simple models suggest. This raises a very important question; is the model which works well in 3D and 2D wrong in lower dimensions, or is there some new physics found in quasi-1D (and to a lesser extent quasi-2D) systems which masks the effect? Perhaps one such possible mechanism can be guessed at immediately, namely the presence of statistical features associated with the small numbers of electrons and impurities in such devices. We shall see that fluctuations are indeed an important facet of transport in extremely small structures.

1.5 Localisation and fluctuation phenomena.

Studies of transport in low dimensional structures have also shown the importance of disorder. Conventional transport theory is based upon the idea that plane wave states (which have a well defined momentum) are capable of carrying current. Scattering causes these plane wave states to scatter into other plane wave states, which controls the size of the current induced by an external stimulus. The scattering is not however considered to affect the nature of the electron states, which remain extended throughout the sample. That things might not be so simple in low dimensional systems became apparent after the pioneering work of Anderson (1958), who showed that if the disorder (i.e. impurity concentration) was strong enough, then the electrons would become localised instead of extended throughout space. Pursuing this idea, Abrahams et al (1979) predicted that all infinite 2D and 1D disordered solids will be insulators at T=0! Even for finite sized systems the effect of this localisation significantly alters the behaviour of electron transport, especially in one dimension. Localisation is a quantum phenomenon which is ignored by conventional semiclassical transport theory, but is clearly an important factor in low dimensional physics. Although localisation effects are also present in bulk specimens, they are (usually) considerably smaller, and so ignoring them does not lead to any noticeable inaccuracies.

In the previous section we touched upon the question of statistics in very small systems. Al'tshuler (1985) and Lee and Stone (1985) have shown that such "mesoscopic" samples exhibit

very interesting fluctuation phenomena and sample specific properties which are not evident in larger samples. For example, the conductance of a given sample will show a reproducible, noise-like structure as a function of chemical potential or magnetic field. Such "Universal Conductance Fluctuations" were initially observed by Umbach et al (1984) in small Au and AuPd rings and wires, and have been reproduced many times since (Webb et al 1987). These fluctuations are washed out as a function of increasing temperature and system size exactly as expected. At T=0, they have universal behaviour which is independent of system shape and only weakly dependent on the dimensionality. Like localisation, these fluctuations are a quantum phenomenon which conventional theories (that calculate average quantities) overlook. Both of these effects, important in small structures and interesting in their own right, will be discussed in detail throughout the rest of this thesis.

1.6 Applications in technology.

With the advent of microfabrication and the increasing degree of miniaturisation in semiconductor technology, electronic components are now reaching a size where quantum confinement effects are becoming important in determining device performance. In addition, and perhaps most excitingly of all, modern growth techniques such as MOCVD (Metal Organic Chemical Vapour Deposition) and MBE (Molecular Beam Epitaxy) have reached a degree of sophistication where it is possible to grow specifically designed structures on the low dimensional length scale (see Physics and Fabrication of Microstructures and Devices ed. Kelly and Wiesbuch 1986). It is then possible to imagine even incorporating the novel physics of quantum confinement into the mode of device operation. These ideas have been explored in the review articles by Board (1985), Kelly and Nicholas (1985), who have discussed for example the superlattice and the quantum well laser, and by Bate (1988). In a series of papers, Capasso et al (1983a,1983b,1984) have discussed compositionally graded devices and pointed out the possibilities for negative differential resistance, switching and storing devices, photodetectors and photomultipliers as well as diodes and transistors of various types. These ideas have been discussed in more detail recently by Heiblum and Eastman (1987). Quantum interference devices based

upon the Aharonov-Bohm effect have also been postulated (Bandyopadhyay et al 1986, Datta and Bandyopadhyay 1987). Already certain of these structures are showing promising signs, and improvement in device performance can only get better as the physics of these ultra-small devices becomes more clearly understood. By studying such devices theoretically we can attempt to see which are the important features, and which are irrelevant as far as device application is concerned.

1.7 Outline of thesis.

In the previous sections we have discussed some of the important aspects of transport in low dimensional (especially quasi-one-dimensional) systems. The rest of this thesis will focus on investigating these and additional features of electrical and thermal transport in such systems, concentrating on quantum confinement effects, localisation and fluctuation phenomena. In chapter 2, a simple Boltzmann method is used to further illustrate the ideas and to calculate the effect of sub-band structure on the transport coefficients in a wire. In chapter 3, the full quantum nature is introduced by considering a more sophisticated formalism based on Green function techniques. The results of Chapter 2 are recalculated which explicitly demonstrates the importance of lifetime broadening in these systems. Chapter 4 is a general review of the theory of localisation and conductance fluctuations relevant to the remaining discussion. Chapter 5 presents a calculation of the weak localisation correction in the quasi-one-dimensional regime, but as a function of an arbitrary number of occupied sub-bands. This allows explicit demonstration of finite width effects ignored in the usual treatments. In Chapter 6, the effects of localisation on the thermal conductivity and thermopower are calculated both in the weak localisation regime and in the vicinity of a Metal-Insulator transition, which leads to a highly plausible scaling theory of thermal transport. Fluctuation effects in thermal transport are also calculated and demonstrated. Finally, in Chapter 7, conclusions are given and possible outlines are presented on the prospects for the future, such as ballistic transport, the relation of fluctuations to 1/f noise, and problems relating to measurement in ultra-small devices.

CHAPTER 2

BOLTZMANN TRANSPORT IN QUASI-ONE-DIMENSIONAL WIRES.

2.1 The transport coefficients.

Under conditions of equilibrium, the conduction electrons in a metal or doped semiconductor are distributed over their energy levels according to the Fermi-Dirac distribution function. Upon applying an external stimulus (an electric field or temperature gradient for example), the electron distribution will be perturbed away from its equilibrium value and, as a result, currents defined on a macroscopic scale will begin to flow. Assuming these stimuli are small, we can relate them linearly to the electric and heat current densities **J** and **Q**,

$$\mathbf{J} = L_{11}\,\mathbf{E}' - \frac{L_{12}}{T}\,\nabla T \tag{2.1a}$$

$$\mathbf{Q} = L_{21}\,\mathbf{E}' - \frac{L_{22}}{T}\,\nabla T \tag{2.1b}$$

where $\mathbf{E}' = \mathbf{E} + |e|^{-1}\,\nabla\mu$, **E** is the applied electric field and μ the chemical potential. The $L_{\alpha\beta}$ are characteristic of the system under consideration and are the quantities we wish to calculate. In general they will be second rank tensors, a complication we shall ignore by restricting the discussion to systems with cubic symmetry where they reduce to scalar multiples of the unit tensor (Butcher 1973). We can then relate the $L_{\alpha\beta}$ to the commonly measured electrical conductivity σ, the thermal conductivity κ and the thermopower S by,

$$\sigma = L_{11} \tag{2.2a}$$

$$\kappa = (L_{22}L_{11} - L_{21}L_{12}) / (TL_{11}) \tag{2.2b}$$

$$S = L_{12} / (TL_{11}). \tag{2.2c}$$

These quantities will be calculated for a quasi-one-dimensional model system when several sub-bands are occupied. Within the framework of Boltzmann theory we shall see that this regime is much more interesting than the one where only the lowest sub-band is occupied.

2.2 Boltzmann transport theory.

Boltzmann theory concerns itself with the calculation of the perturbed electron distribution function. Although it treats the underlying dynamics very simply, it nevertheless provides intuitive insight into an otherwise complicated problem (Greenwood 1958), and has proved very

successful in analysing a wide range of phenomena in bulk semiconductors and metals (Blatt 1968, Nag 1980, Butcher 1986). Since the method is well known, we shall content ourselves with only a brief description. The central idea is that since electrons cannot be destroyed, the perturbed distribution function $f(\mathbf{k},\mathbf{r},t)$ must obey a continuity equation of the following type,

$$\frac{\partial f}{\partial t} + \mathbf{v}.\nabla_{\mathbf{r}} f - \frac{|e|}{\hbar}\mathbf{E}'.\nabla_{\mathbf{k}} f = \left[\frac{\partial f}{\partial t}\right]_{coll} \tag{2.3}$$

Equation (2.3) represents the competition between the applied fields which tend to drive the system away from equilibrium and scattering (contained within the collision term) which tends to restore it. Since impurity scattering dominates at low temperatures (where most of the interesting effects are observed) we shall consider this particular mechanism in detail. The word impurity will be used somewhat loosely, it could refer to donor atoms for example or equally well to defects and disorder within the solid. All such static deviations from periodicity will scatter the electrons elastically, which further simplifies the analysis. Once we have calculated the distribution function for a particular model, the electric and heat current densities are given by (Butcher 1973, 1986),

$$\mathbf{J} = -\frac{|e|}{\pi}\sum_{\alpha}\int \mathbf{v}_{\alpha} f_{\alpha}\, d\mathbf{k}' \tag{2.4a}$$

$$\mathbf{Q} = \frac{1}{\pi}\sum_{\alpha}\int \mathbf{v}_{\alpha}(\varepsilon-\mu) f_{\alpha}\, d\mathbf{k}' \tag{2.4b}$$

We shall also assume that steady state has been attained, and leave it understood that if f is an explicit function of the variable $\mathbf{r}$, then equations (2.4a) and (2.4b) represent the current densities at the point $\mathbf{r}$. The index α is a band index, and in the present problem refers to the sub-bands.

2.3 The quasi-one-dimensional regime.

In a quasi-1D system, the electrons in the $(m,n)th$ sub-band will be described by a distribution function $f_{mn}(k,\mathbf{r})$, each of which will obey an equation of the type (2.3). The collision term must account for scattering between the sub-bands as well as scattering within the same sub-band, and so we write,

$$\left[\frac{\partial f}{\partial t}\right]_{coll} = \sum_{ij}\int [P(k'ij;kmn) f_{ij}(k')(1-f_{mn}(k)) \tag{2.5}$$

$$- P(kmn;k'ij) f_{mn}(k)(1-f_{ij}(k'))]\,\frac{L}{2\pi}\,dk'$$

where $P(kmn;k'ij)$ is the transition rate from a full state $|kmn>$ to an empty state $|k'ij>$. The first term accounts for scattering into the state $|kmn>$ and the second for scattering out of it. Equations (2.3) and (2.5) then provide a set of coupled equations for the distribution functions. To solve them we make a number of approximations of a standard type (Butcher 1973, Nag 1980). Firstly we note that since the scattering is assumed to be elastic, detailed balance implies,

$$P(kmn;k'ij) = P(k'ij;kmn) \tag{2.6}$$

Secondly, by assuming the fields are weak we can linearise f by writing,

$$f_{mn} = f^0 + f^1_{mn} \tag{2.7}$$

where f^0 is the Fermi-Dirac distribution and f^1_{mn} is a small correction which is linearly related to $\mathbf{E}'$ and ∇T. Thirdly, we can employ the relaxation time approximation (strictly valid for elastic scattering and symmetric energy bands),

$$\left[\frac{\partial f}{\partial t}\right]_{coll} = -\frac{f^1_{mn}}{\tau_{mn}} \tag{2.8}$$

τ_{mn} is an energy dependent relaxation time which characterises the electron-impurity interaction. Within these three approximations we find,

$$f^1_{mn} = \tau_{mn}\, v_{mn} \left[|e|E' + \frac{1}{T}\frac{dT}{dz}(\varepsilon-\mu)\right]\frac{\partial f^0}{\partial \varepsilon} \tag{2.9}$$

v_{mn} is the velocity of an electron in sub-band (m,n) with wavevector k. We are going to assume that the sub-bands are parabolic, whereupon $v_{mn} = \hbar k/m^* = (2(\varepsilon-\varepsilon_{mn})/m^*)^{1/2}$. It is of course possible to include non-parabolic effects (Nelson et al 1987), but we expect the effect on the transport coefficients to be small unless the non-parabolicity in unusually large. The relaxation times are given by,

$$\frac{1}{\tau_{mn}} = \sum_{i,j}\int P(kmn;k'ij)\frac{L}{2\pi}\,dk' - \sum_{i,j}\int P(k'ij;kmn)\frac{\tau_{ij}}{\tau_{mn}}\frac{k'}{k}\frac{L}{2\pi}\,dk'. \tag{2.10}$$

There are several points which should be stressed about the relaxation times. In the case of multi-sub-band transport they are generally coupled together because of inter-sub-band scattering (Siggia and Kwok 1970, Milsom and Butcher 1986). Secondly, the relaxation times are not the scattering times of plane wave states (we shall return to this shortly). Finally, they are only independent of temperature because we have assumed the scattering to be entirely elastic. There are various subtleties associated with elastic scattering which must be handled carefully, and we

shall address them as they are encountered. In the presence of inelastic scattering equation (2.10) would explicitly contain the Fermi function.

The relaxation times may be calculated once the transition rate is specified. For this we use Fermi's Golden Rule:

$$P(kmn;k'ij) = \frac{2\pi}{\hbar} |\langle kmn | V | k'ij \rangle|^2 \delta(\varepsilon-\varepsilon') \tag{2.11}$$

which requires the form of the wavefunctions in the absence of disorder and the impurity potential to be known. Clearly for a qualitative discussion of transport through a device, the exact form of the sub-band wavefunctions is less important than the fact that there are discontinuities in the density of states (although this is not necessarily true for other quantities such as optical absorption for example). We therefore content ourselves with a model confining potential of the infinite "box" type where,

$$\phi_{mn}(k) = \left[\frac{4}{Lab}\right]^{\frac{1}{2}} e^{ikz} \sin(n\pi x/a) \sin(m\pi y/b) \tag{2.12}$$

and a and b are the channel widths in the x and y directions of a wire of length L ($L \gg a,b$). More sophisticated calculations have been performed (Berggren and Newson 1986, Laux et al 1988) and show this basis set is a reasonable one. The simplest form of scattering which is elastic and pointlike is of the standard δ – function type,

$$V(\mathbf{r}) = \sum_l U_0 \, \delta(\mathbf{r}-\mathbf{r}_l) \tag{2.13}$$

where $\mathbf{r}_l$ is the position of the *lth* impurity and U_0 is a measure of the strength of the interaction. δ – function scatterers have the nice, additional feature of being velocity randomising in the sense that scattering to another sub-band is equally favoured to states k and $-k$ in that sub-band. The last term in equation (2.10) is then zero and the equations decouple (a similar effect occurs in 3D systems and is well known (Butcher 1973)). The transition rate however still depends upon the positions of all the impurities. This dependence may be removed by assuming the system is sufficiently large and the impurity coordinates sufficiently uncorrelated that we can average over all possible impurity configurations. This ensemble averaging was first discussed in detail by Kohn and Luttinger (1957). The fundamental assumption behind it is that the mean behaviour of all macroscopically similar samples is the same as the typical behaviour of a given sample.

Naturally this neglects fluctuation effects such as Universal Conductance Fluctuations (Lee et al 1987) and we will return to discuss these later. Configuration averaging (by integrating over the impurity coordinates) assuming the density of impurities is small gives,

$$\frac{1}{\tau_{mn}(\varepsilon)} = 2n_0\left[\frac{U_0}{2\hbar ab}\right]^2\left[\frac{m^*}{2}\right]^{1/2}\sum_{i,j}\frac{(2+\delta_{im})\,(2+\delta_{jn})\,\theta(\varepsilon-\varepsilon_{ij})}{(\varepsilon-\varepsilon_{ij})^{1/2}} \qquad (2.14)$$

where n_0 is the number of scatterers per unit length. It should be noted that in the limit a or b tending to zero the right hand side of this expression appears to diverge. This is an artifact of δ – function potentials; for a potential with a finite range R the relaxation times become independent of a and b in the limit $a \ll R$, $b \ll R$ as may easily be verified. There is no problem however in discussing the physical behaviour of the transport coefficients by using zero-range scattering potentials.

A schematic plot of the relaxation times as a function of energy in the lowest three sub-bands is shown in figure 2.1. The behaviour one would associate with a purely 1D system, namely $\tau \sim \varepsilon^{1/2}$ is clearly visible in the lowest sub-band. When ε coincides with a sub-band minimum however, the relaxation times drop discontinuously to zero. This reflects the form of the quasi-1D density of states (figure 1.5), since at each new sub-band minimum there is an infinite density of states for the electrons to scatter to. This structure in $\tau(\varepsilon)$ is more pronounced than that found in 2D structures (see e.g. Mori and Ando 1979, Sernelius et al 1985)), precisely because the density of states shows sharper discontinuities. Since the transport properties of the system depend upon $\tau(\varepsilon)$, we should therefore expect to see considerable structure in the transport coefficients as the chemical potential is varied. This structure will be evident in the vicinity of the sub-band minima, irrespective of the actual value of the energies of these minima, which is another reason why the exact sub-band wavefunctions do not need to be known.

2.4 The electrical conductivity.

Having found the perturbed distribution function, we can insert it into equations (2.4a) and (2.4b) to derive expressions for the electric and heat currents. Comparison with equations (2.1a) and (2.1b) directly leads to the $L_{\alpha\beta}$ and hence the electrical conductivity, thermal conductivity and thermopower. We find,

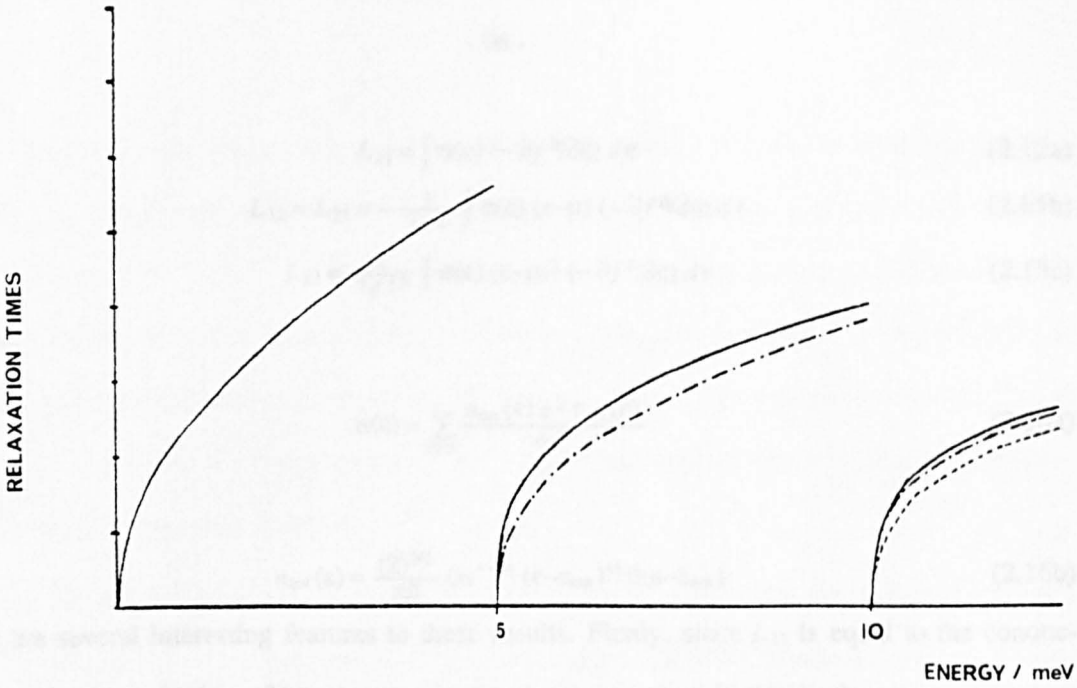


Figure 2.1: A schematic plot of the relaxation times of the lowest three sub-bands as a function of energy. The second sub-band minimum is at 5 meV, the third is at 10 meV.

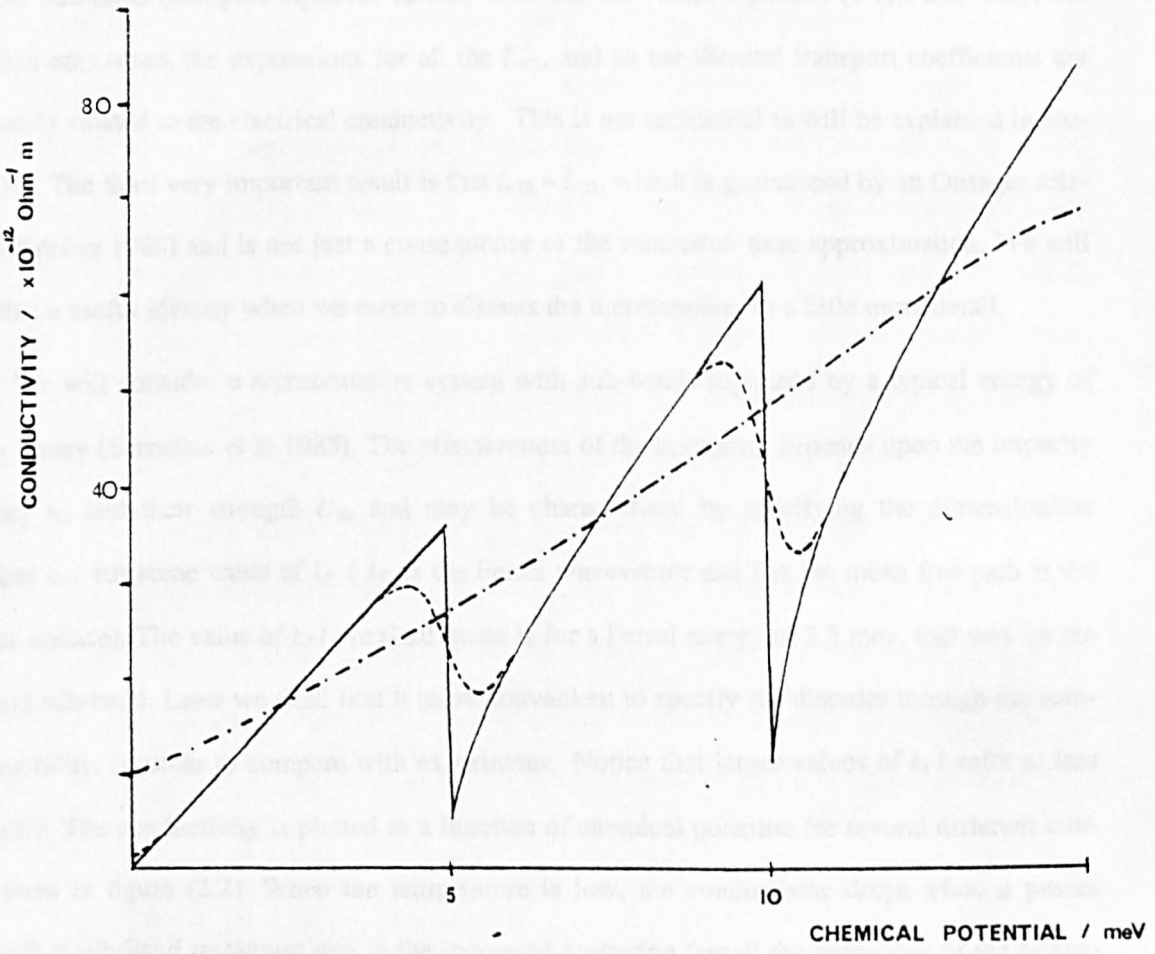


Figure 2.2: The conductivity as a function of chemical potential for three different temperatures and $k_F l = 30$: ——— $T = 0.3K$, ------ $T = 3K$ and -·-·-· $T = 30K$.

$$L_{11} = \int \sigma(\varepsilon)\,(-\partial f^0/\partial\varepsilon)\,d\varepsilon \tag{2.15a}$$

$$L_{12} = L_{21} = -\frac{1}{|e|}\int \sigma(\varepsilon)\,(\varepsilon-\mu)\,(-\partial f^0/\partial\varepsilon)\,d\varepsilon \tag{2.15b}$$

$$L_{22} = \frac{1}{|e|^2}\int \sigma(\varepsilon)\,(\varepsilon-\mu)^2\,(-\partial f^0/\partial\varepsilon)\,d\varepsilon \tag{2.15c}$$

where,

$$\sigma(\varepsilon) = \sum_{m,n}\frac{n_{mn}(\varepsilon)\,e^2\,\tau_{mn}(\varepsilon)}{m^*} \tag{2.16a}$$

and,

$$n_{mn}(\varepsilon) = \frac{(2)^{3/2}}{\pi\hbar^3}\,(m^*)^{3/2}\,(\varepsilon-\varepsilon_{mn})^{1/2}\,\theta(\varepsilon-\varepsilon_{mn}). \tag{2.16b}$$

There are several interesting features to these results. Firstly, since L_{11} is equal to the conductivity, and at $T=0$, $\partial f/\partial\varepsilon = -\delta(\varepsilon-\varepsilon_F)$, $\sigma(\varepsilon_F)$ in the above equations is simply the zero temperature conductivity of the system. Correspondingly, $n_{mn}(\varepsilon_F)$ is simply the density of electrons in the $(m,n)th$ sub-band (compare equation (2.16a) with the 3D result, equation (1.1)). Secondly, the function $\sigma(\varepsilon)$ enters the expressions for all the $L_{\alpha\beta}$, and so the thermal transport coefficients are intimately related to the electrical conductivity. This is not accidental as will be explained in section 3.5. The third very important result is that $L_{12} = L_{21}$, which is guaranteed by an Onsager relation (Butcher 1986) and is not just a consequence of the relaxation time approximation. We will find this a useful identity when we come to discuss the thermopower in a little more detail.

We will consider a representative system with sub-bands separated by a typical energy of order 5 mev (Sernelius et al 1985). The effectiveness of the scattering depends upon the impurity density n_0 and their strength U_0, and may be characterised by specifying the dimensionless number $k_F l$ for some value of k_F (k_F is the Fermi wavevector and l is the mean free path at the Fermi surface). The value of $k_F l$ we shall quote is for a Fermi energy of 2.5 mev, half way up the ground sub-band. Later we shall find it more convenient to specify the disorder through the sample mobility, in order to compare with experiment. Notice that larger values of $k_F l$ refer to less disorder. The conductivity is plotted as a function of chemical potential for several different temperatures in figure (2.2). When the temperature is low, the conductivity drops when μ passes through a sub-band minimum due to the increased scattering (recall the behaviour of the relaxation times). Such features in the transport coefficients are known collectively as quantum size

effects or QSE. As the temperature is raised, thermal broadening becomes more important until when $k_B T$ is comparable to the sub-band separation all the structure is smeared out (5 mev corresponds to approximately 30 K). In this regime, σ is almost linear in μ and all information about the sub-bands is lost. This explains incidently why experiments have to be performed at low temperatures if QSE are to be observed. We have not plotted the conductivity as a function of disorder ($k_F l$), since in Boltzmann theory σ is simply proportional to $k_F l$ or inversely proportional to the impurity concentration.

The above calculation is very simplistic and supports our conjecture that sub-band structure should lead to QSE in the conductivity. This idea is not new, and has been anticipated in 2D systems for a long time (Mori and Ando 1979, Sernelius et al 1985, Milsom 1987). Inclusion of inelastic scattering via phonons for example does little to change the structure of the results (Fishman 1987, Milsom and Butcher 1986). Comparison with experiments produces in general, however, very poor agreement. The most suitable device for performing the experiments on is the pinched MOSFET (see introductory chapter), since it allows the chemical potential to be varied continuously. In the experiments carried out by Skocpol et al (1984), some tantalising evidence for QSE was observed but was very inconclusive. Further experiments on devices with greater resistance (and hence more disorder) (Hartstein et al 1984) produced even less agreement, with marked (though reproducible) fluctuations being apparent. In an effort to reduce the effect of these fluctuations (to try and resolve sub-band structure in the conductivity), Warren et al (1986) performed an experiment on 250 parallel inversion lines simultaneously, and produced perhaps the best evidence for QSE in quasi-1D conduction. What is very clear however is that the size of the effect is very much reduced from the Boltzmann predicted values, and that the role of increasing disorder is much more important than this simple theory would have us believe. To treat quasi-one-dimensional systems properly we are therefore going to have to go beyond simple semiclassical theories.

2.5 Thermal transport in sub-band systems.

It is useful to calculate the thermal transport coefficients within the Boltzmann framework

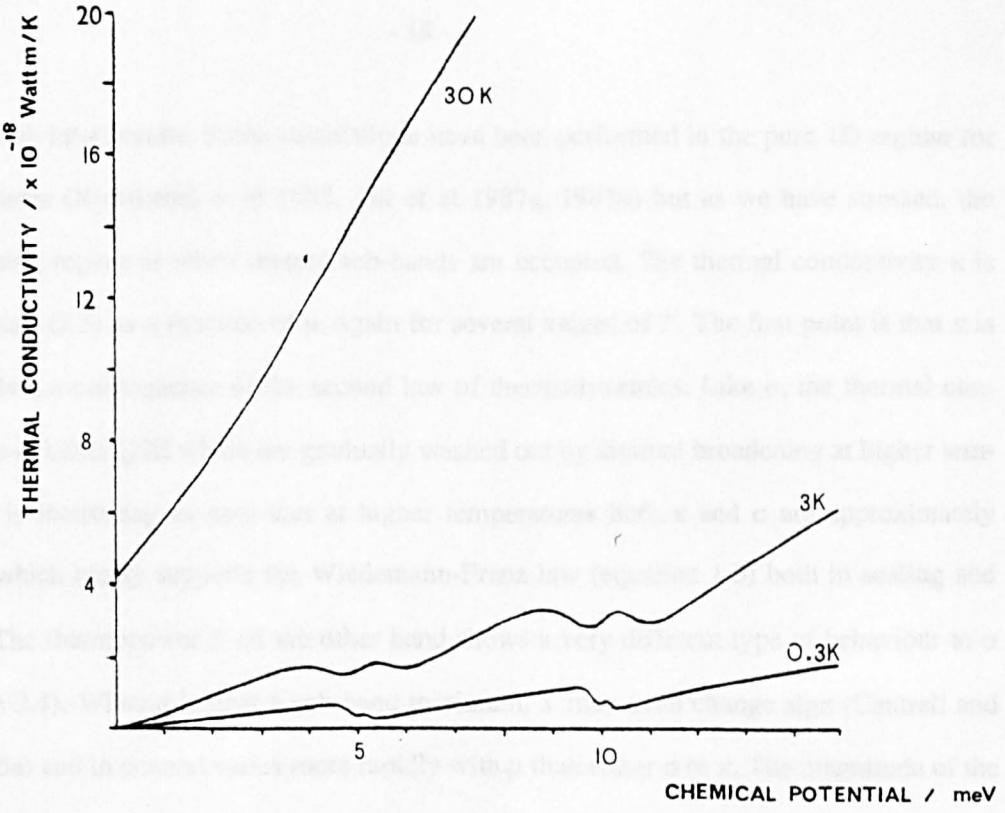


Figure 2.3: The thermal conductivity of the same system for the same three temperatures as in the previous figure.

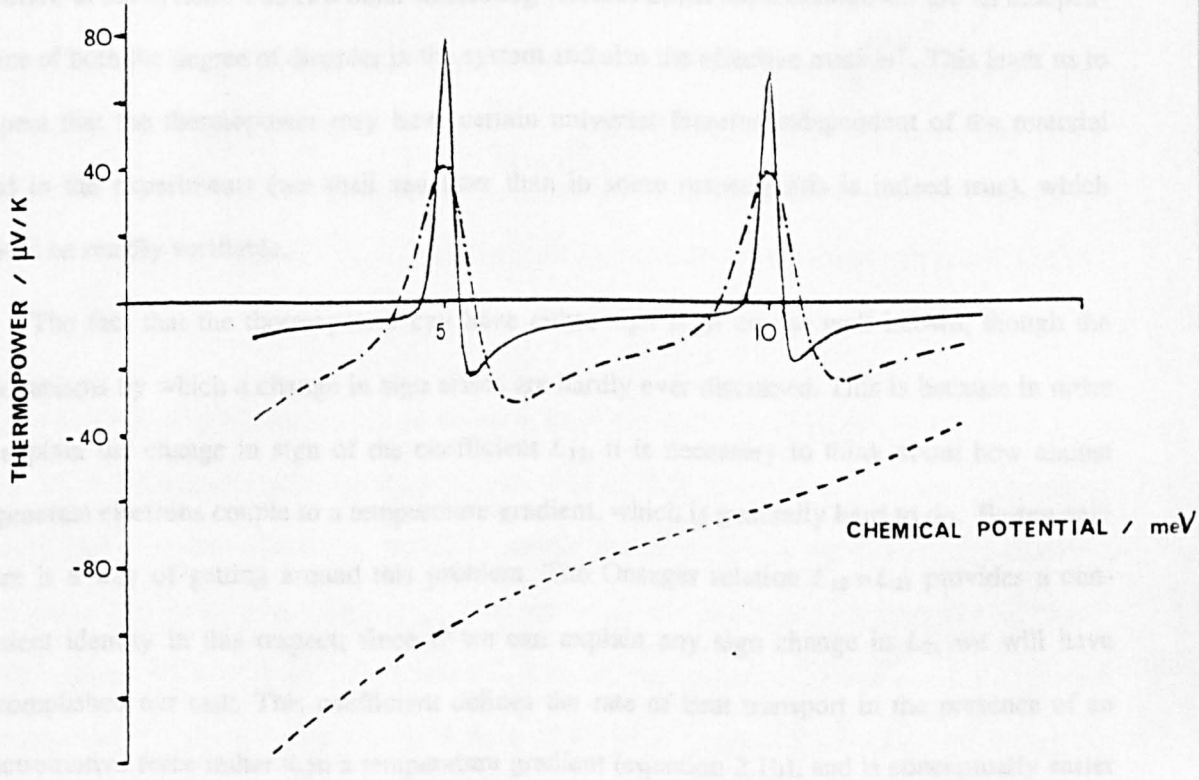


Figure 2.4: The thermopower as a function of chemical potential for, ——— $T = 0.3K$, -·-·- $T = 3K$ and ····· $T = 30K$.

to compare with later results. Some calculations have been performed in the pure 1D regime for these coefficients (Kubakaddi et al 1985, Jali et al 1987a, 1987b) but as we have stressed, the more interesting regime is when several sub-bands are occupied. The thermal conductivity κ is plotted in figure (2.3) as a function of μ, again for several values of T. The first point is that κ is always positive, a consequence of the second law of thermodynamics. Like σ, the thermal conductivity also exhibits QSE which are gradually washed out by thermal broadening at higher temperatures. It is interesting to note that at higher temperatures both κ and σ are approximately linear in μ, which nicely supports the Wiedemann-Franz law (equation 1.3) both in scaling and magnitude. The thermopower S on the other hand shows a very different type of behaviour to σ and κ (figure 2.4). When μ is near a sub-band minimum, S may even change sign (Cantrell and Butcher 1985a) and in general varies more rapidly with μ than either σ or κ. The magnitude of the positive peaks is also rather large (of order 70 μVK^{-1}). This suggests that if we wish to observe QSE, then the thermopower is the best quantity to measure being so sensitive to the sub-band structure in the system. The two other interesting features about the thermopower are its independence of both the degree of disorder in the system and also the effective mass m^*. This leads us to suspect that the thermopower may have certain universal features independent of the material used in the experiments (we shall see later than in some respects this is indeed true), which should be readily verifiable.

The fact that the thermopower can have either sign is of course well known, though the mechanisms by which a change in sign arises are hardly ever discussed. This is because in order to explain the change in sign of the coefficient L_{12}, it is necessary to think about how almost degenerate electrons couple to a temperature gradient, which is generally hard to do. Fortunately there is a way of getting around this problem. The Onsager relation $L_{12}=L_{21}$ provides a convenient identity in this respect, since if we can explain any sign change in L_{21} we will have accomplished our task. This coefficient defines the rate of heat transport in the presence of an electromotive force rather than a temperature gradient (equation 2.1b), and is conceptually easier to think about. Consider a Fermi sea of electrons. At finite temperature a certain number will be excited above the Fermi surface, leaving some hole-like excitations below it. When an electric

field is applied, the electrons and holes drift in opposite directions but since they have opposite charge, they both contribute to the current in the same sense. The conductivity is thus always positive. When we talk about heat transfer however, the electrons and holes carry heat in opposite directions. When $k_B T$ is small these two contributions nearly cancel. Normally, the heat flux due to the electron-like excitations will be slightly greater than the heat flux transported by the holes (in the opposite direction) because the electrons have slightly greater velocity. If the Fermi level lies just below a sub-band minimum however, the electrons will experience a rapid increase in scattering compared to the holes because their excited energies lie in the region of this minimum. They therefore transport heat far less effectively and the hole contribution takes over. The coefficient L_{21} changes sign in this case, and as a result, so does the thermopower. Having seen this argument we can now understand why the coefficient L_{12} might change sign. When a temperature gradient is applied (in the absence of any electric fields), the electrons and holes both drift in the same direction (from hot to cold) and the two contributions to the electric current nearly cancel. The argument for the sign change then goes through exactly as before.

To see how long these positive peaks stay in the thermopower, we should calculate $S(\mu,T)$ as a function of temperature rather than chemical potential. When doing this, it should be remembered that μ is also a function of T, because the number of electrons in the system remains constant. This is often overlooked. The variation $\mu(T)$ may therefore be calculated numerically from,

$$N = \frac{L}{\pi}\left[\frac{2m^*}{\hbar^2}\right]^{1/2} \sum_{m,n} \int_{\varepsilon_{mn}}^{\infty} \frac{f^0(\varepsilon)}{(\varepsilon-\varepsilon_{mn})^{1/2}}\, d\varepsilon \qquad (2.17)$$

where the number of electrons N is determined from the Fermi energy ($\varepsilon_F = \mu(0)$) in the limit $T \to 0$. Several results for $\mu(T)$ are shown in figure (2.5). The variation of μ is most pronounced when ε_F lies just below a sub-band minimum; the proximity of the next sub-band means there is a large number of states with ε slightly greater than ε_F for the electrons to occupy as T is raised, which makes μ fall off rapidly. Taking the temperature dependence of μ into account, S is plotted as a function of T for two values of ε_F in figure (2.6). Near a sub-band minimum we can expect positive thermopower to persist up to $T<6K$. When ε_F is many $k_B T$ away from a sub-band minimum , S is negative and linear in T, and is well represented by the usual Mott formula (equa-

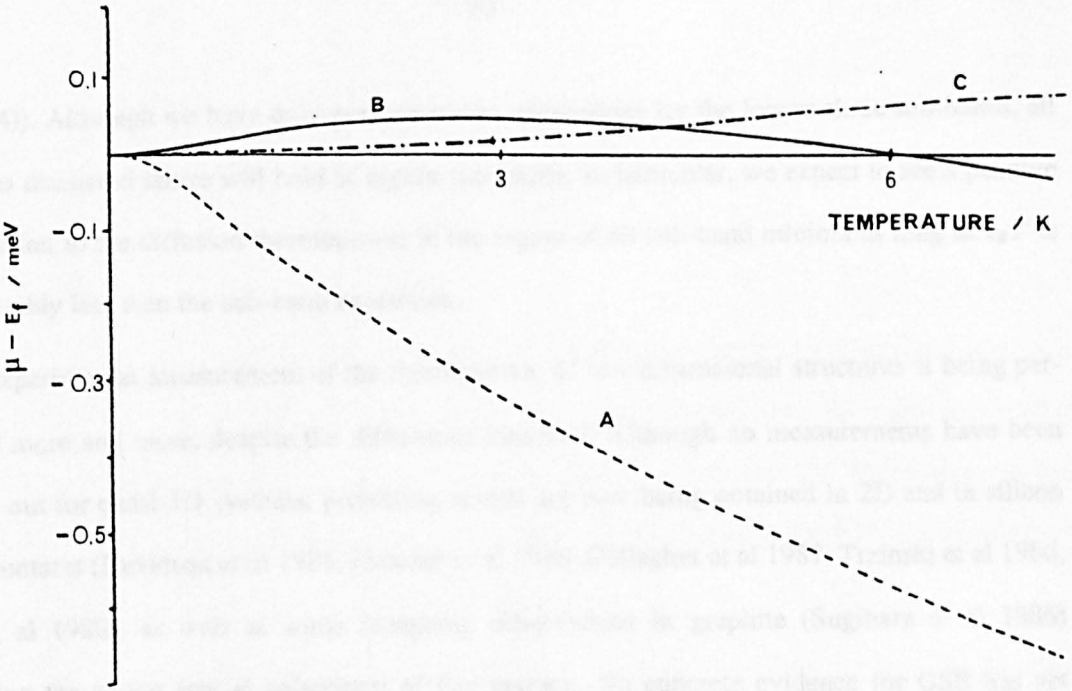


Figure 2.5: The deviation of the chemical potential from the Fermi energy as a function of temperature, for three different values of the Fermi energy ε_F A) 4.9 meV, B) 5.2 meV and C) 7.5 meV. The second sub-band minimum is at 5 meV.

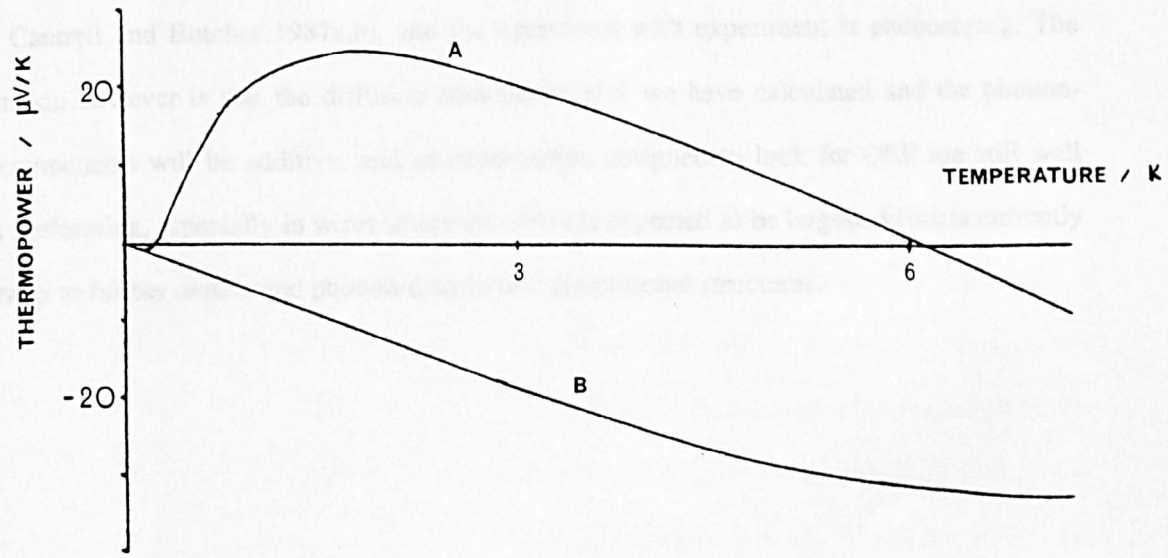


Figure 2.6: The thermopower as a function of temperature for two different values of the Fermi energy A) 4.9 meV and B) 7.5 meV, and $k_F l = 30$.

tion (1.4)). Although we have only performed the calculations for the lowest three sub-bands, all the ideas discussed above will hold in higher sub-bands. In particular, we expect to see a positive contribution to the diffusion thermopower in the region of all sub-band minima as long as $k_B T$ is considerably less than the sub-band separation.

Experimental measurement of the thermopower of low dimensional structures is being performed more and more, despite the difficulties involved. Although no measurements have been carried out for quasi-1D systems, promising results are now being obtained in 2D and in silicon microcontacts (Davidson et al 1986, Fletcher et al 1986, Gallagher et al 1987, Trzinski et al 1986, Ruf et al 1988), as well as some intriguing observations in graphite (Sugihara et al 1986) reflecting the strong spatial anisotropy of this system. No concrete evidence for QSE has yet been universally found (the experiments so far have not generally been looking for them), though very recent experiments have shown a change of sign of the type expected (C.Ruf private communication). It has been demonstrated, however, that phonon-drag (which we have ignored) is also important. (In the experiments of Ruf et al they subtracted this phonon-drag component from the total result in order to show a sign change in the diffusive component of the thermopower). Simple calculations for the phonon-drag thermopower in $2D$ have been performed (Nicholas 1985, Cantrell and Butcher 1987a,b), and the agreement with experiment is encouraging. The expectation however is that the diffusive component of S we have calculated and the phonon-drag components will be additive, and so experiments designed to look for QSE are still well worth performing, especially in wires where the effect is expected to be largest. Work is currently underway to further understand phonon-drag in low dimensional structures.

CHAPTER 3

LIFETIME BROADENING OF SUB-BAND STRUCTURE.

3.1 Introduction to lifetime broadening.

We have discussed Boltzmann transport theory in a quasi-one-dimensional wire, and found that the results for the conductivity do not agree well with experiment. We therefore need to improve the theory and go beyond the simple approximations that the Boltzmann method assumes. Hopefully we will find that some of the features highlighted in the previous chapter will persist and surface also in a more sophisticated treatment.

Certainly one approximation that Boltzmann theory relies on, namely that the scattering does not affect the single particle properties such as the nature of the electron states between collisions or the density of states, must be considered dubious. On general quantum mechanical grounds, a state characterised by a lifetime τ has a natural uncertainty in its energy $\sim\hbar/\tau$, and this broadening must modify the single particle properties. For weakly disordered 3D systems this effect is usually small, and correspondingly the Boltzmann method often works quite well. For low dimensional systems on the other hand this need not be the case at all (Cantrell and Butcher 1985b). One consequence of this will be that the QSE in the transport coefficients are considerably reduced when broadening is introduced into the theory.

Our aim then is to develop a systematic theory which may be applied to a quasi-one-dimensional wire and which implicitly allows for broadening. This is achieved using Green's function methods and standard field theory techniques. Unlike most other workers we choose to work at finite temperature at the outset through the Matsubara method. The results obtained will also be of considerable use later in this work. An approximate calculation will then be performed and for the purposes of comparison, the results of the last chapter are recalculated. For general accounts of the methods used the reader is referred to standard texts e.g. Abrikosov et al 1965, Fetter and Walecka 1971 or Mahan 1981. The book by Nakajima et al 1980 provides an excellent introduction to the ideas of Fermi-Liquid theory, and at a simpler level, an illuminating discussion of the impurity problem is given by Doniach and Sondheimer 1974. The specific

question of doping effects on the electron states in semiconductors has been addressed by Bonch-Bruevich 1966.

3.2 The single-particle Green's function.

The model used previously for the unperturbed system provides a convenient set of basis states for the calculation. Introducing the field operators,

$$\psi(\mathbf{x}) = \sum_{\alpha,k} \phi_{\alpha k}(\mathbf{x})\, C_{\alpha k} \tag{3.1a}$$

$$\psi^{+}(\mathbf{x}) = \sum_{\alpha,k} \phi^{*}_{\alpha k}(\mathbf{x})\, C^{+}_{\alpha k} \tag{3.1b}$$

the N-electron Hamiltonian may be cast into second quantised form,

$$H = \frac{1}{L}\sum_{\alpha k} (\hbar^2 k^2/2m^{*} + \varepsilon_{\alpha})\, C^{+}_{\alpha k}\, C_{\alpha k} + \frac{1}{L}\sum_{\alpha,\beta,q,l} M_{\alpha\beta}(X_l, Y_l, q)\, e^{iqz_l}\, C^{+}_{\alpha,k+q}\, C_{\beta,k} \tag{3.2}$$

$$M_{\alpha\beta}(X_l, Y_l, q) = \int \xi^{*}_{\alpha}\, U(x - X_l, y - Y_l, q)\, \xi_{\beta}\, dx\, dy$$

with U the potential of the lth scatterer and q the wavenumber of the Fourier transform of the z coordinate. For convenience the sub-band labels are labelled by one index α, and the spin labels suppressed. The single particle Green's function which describes the excitations at finite temperatures is,

$$G_{\alpha\beta}(k, k', \tau) = -\langle T_{\tau}\, C_{\beta k'}(\tau)\, C^{+}_{\alpha k}(0)\rangle \tag{3.3}$$

where $\langle \ldots. \rangle$ denotes the thermodynamic average over the N-particle states of the system. (Further discussions and details may be found in Appendix A). Equation (3.3) may be Fourier transformed to give a more useful representation (Mahan 1981),

$$G_{\alpha\beta}(k, k', ip_n) = \int_0^{\beta} d\tau\, G_{\alpha\beta}(k, k', \tau)\, e^{i\tau p_n}. \tag{3.4}$$

The Matsubara frequencies for Fermions are odd; $p_n = (2n + 1)\pi/\beta$ with $\beta = 1/k_B T$, and contain information about the temperature of the system.

The (real-time) retarded Green's function may be obtained from the Matsubara function by making the analytic continuation to the real axis $ip_n \rightarrow \varepsilon + i\delta$ (see Appendix A). The variable ε is an energy variable with domain $-\infty < \varepsilon < \infty$, and is measured relative to the natural energy origin which is the chemical potential. As in a general Fermi-Liquid theory, ε describes the energy of an excitation in the system; those with $\varepsilon > 0$ are then quasi-particles and those with $\varepsilon < 0$ quasi-holes.

Many useful properties of the system are given by the retarded (and advanced) Green's functions. For example, the single-particle density of states is related to the diagonal components of G^R (see Appendix A),

$$g(\varepsilon) = -\frac{1}{\pi} \sum_{\alpha,k} \mathrm{Im}\, G^R_{\alpha\alpha}(k, k, \varepsilon) \tag{3.5}$$

Another quantity of interest is the spectral function $A_\alpha(k,\varepsilon)$ defined by,

$$A_\alpha(k,\varepsilon) = -2\, \mathrm{Im}\, G^R_{\alpha\alpha}(k, k, \varepsilon) \tag{3.6}$$

which is interpreted as being the probability that an excitation labelled by $|k, \alpha\rangle$ has energy ε. It thus has an associated conservation law,

$$\int_{-\infty}^{\infty} \frac{d\varepsilon}{2\pi} A_\alpha(k,\varepsilon)\, d\varepsilon = 1 \tag{3.7}$$

which is also shown in Appendix A. Not only are G^R and G^A useful in determining the equilibrium properties, we shall see that they are also related to the transport coefficients in the linear response regime.

The second term in the Hamiltonian (3.2) represents the perturbation due to the impurities. To evaluate the Green's function systematically in the presence of this perturbation, we write it in the form,

$$G_{\alpha\beta}(k, k', \tau) = -\frac{\langle T_\tau [S(\beta)\, C_{\beta k'}(\tau)\, C^+_{\alpha k}(0)]\rangle_0}{\langle S(\beta)\rangle_0} \tag{3.8}$$

where $\langle \ldots. \rangle_0$ now denotes a thermodynamic average over the unperturbed states of the system, and the operators are defined in the interaction representation. Expanding the $S-$ matrix, $S(\beta)$, produces an infinite number of terms which may be conveniently represented graphically using Feynman diagrams. The important diagrams are those which when summed, yield the proper self-energy (Fetter and Walecka 1971, Mahan 1981) which is related to the Green's function via a Dyson equation. The self-energy Σ has several important mathematical properties which hold irrespective of the form of the perturbation (Nakajima et al 1980). If we regard $\Sigma_{\alpha\beta}(k, k', Z)$ as a function of the complex variable Z, then Σ is analytic except on the real axis. Furthermore, the real axis is a branch cut of the function such that the analytic continuation $Z \to x \pm i\delta$ (x real) yields,

$$\Sigma(Z \to x \pm i\delta) = \Delta(x) \mp i\Gamma(x) \quad , \Gamma(x)>0 \tag{3.9}$$

where the real part $\Delta(x)$ shifts the energies of the excitations and the imaginary part $\Gamma(x)$ determines their damping. The imaginary part is the most important in determining the transport properties of a given system because it directly relates to the scattering within the system.

The self-energy is configuration averaged to remove the dependence of the impurity coordinates (Kohn and Luttinger 1957, Edwards 1958). For a 3D system on a macroscopic length scale, this implies the system is homogeneous (and as such translationally invariant), whereupon the self energy and hence Green's function will be diagonal in **k** (Abrikosov et al 1965). For a subband system, impurity averaging still leads to Σ and G being diagonal in **k**, but not in the subband indices (because of the lack of translational symmetry in the confined directions). As a result, the Dyson equation in general is of the matrix type,

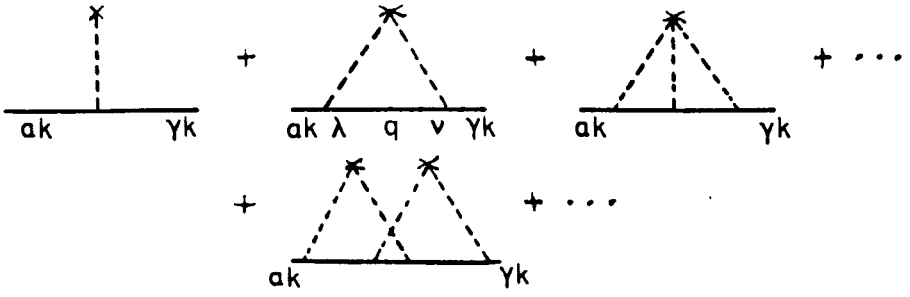
$$G_{\alpha\beta}(k,ip_n) = G^0_\alpha(k,ip_n)\delta_{\alpha\beta} + \sum_\gamma G^0_\alpha(k,ip_n)\,\Sigma_{\alpha\gamma}(k,ip_n)\,G_{\gamma\beta}(k,ip_n) \tag{3.10}$$

where G^0_α is the unperturbed Matsubara function,

$$G^0_\alpha(k,ip_n) = \frac{1}{ip_n - \xi_{\alpha k}} \tag{3.11}$$

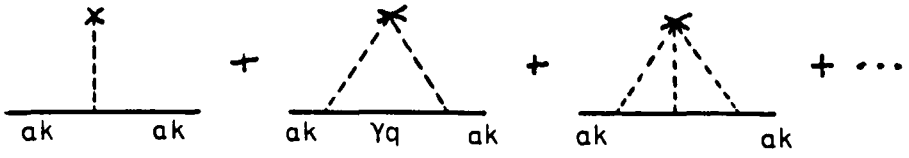
$$\xi_{\alpha k} = \hbar^2 k^2/2m^* + \varepsilon_\alpha - \mu$$

For impurity scattering, the proper self-energy has contributions of the following type after configuration averaging (Mahan 1981),



A cross represents scattering from a single scatterer, a dotted line denotes an electron-impurity interaction which carries the appropriate matrix element and a full line is the full (perturbed) Green's function. Since the scattering is taken to be elastic, all the Green's functions in the diagrams have the same frequency ip_n, and so when evaluating the self-energy there are no frequencies summations to be performed. In principle we can now sum all the self-energy diagrams and derive the exact Green's function through equation (3.10). This is of course impractical, and

so for the purposes of calculation we must retain those self-energy diagrams which are most important and disregard the rest. To this end we are going to assume that the impurity concentration is low and that the sub-bands are well separated in energy. Diagrams containing more than one cross are therefore of lower order in the impurity concentration than those with a single cross, and may be ignored. The advantage of discussing the case where the sub-bands are well separated is that the off-diagonal elements are then smaller than the diagonal elements and may also be ignored (Cantrell and Butcher 1985, Takeshima 1986). The remaining terms then look like,



The first diagram is simply a constant which shifts the energy origin and is trivial. The second diagram involving one internal Green's function is the first non-trivial contribution, and retaining solely this term is called the self-consistent Born approximation. Actually all the single cross diagrams in this series may be summed by invoking a $T-$ matrix (Mahan 1981), however we refrain from doing this because the final answer (for this model) constitutes little more than a trivial renormalisation of the self-consistent Born approximation. We shall therefore limit the discussion to working to this order.

When performing calculations of this type, it is often found to be sufficient to use the unperturbed Green's function instead of the perturbed function in the self-energy diagrams (Abrikosov et al 1965, Edwards 1958, Langer 1960). If we do that in the present problem however, we find that the self-energy contains singularities of the type $(\varepsilon - \xi_{ak})^{-1/2}$, which are related to the singularities in the unperturbed density of states and are a serious problem when carrying out calculations in quasi-one-dimensional systems. This problem does not arise in quasi-2D or 3D problems (compare the relative forms of the densities of states which have no singularities). It is therefore essential when doing calculations of this type to use the self-consistent approximation which removes these divergences. The physical interpretation behind this approximation is clear;

between collisions electrons are not free but move in the background potential of all the other scatterers and are consequently damped (Langer and Neal 1966). The Green's function and self-energy must then be solved self-consistently retaining the same degree of approximation for both. Within all these approximations we then have,

$$G_{\alpha\alpha} = \frac{1}{\varepsilon - \xi_{\alpha k} - \Sigma_{\alpha\alpha}(k, ip_n)} \tag{3.12a}$$

$$\Sigma_{\alpha\alpha}(k, ip_n) = \frac{n_0}{L} \sum_{\beta q} \frac{<M_{\alpha\beta} M_{\beta\alpha}>}{\varepsilon - \varepsilon_\beta - \hbar^2(k-q)^2/2m^* - \Sigma_{\beta\beta}(k, ip_n)} \tag{3.12b}$$

where < .... > means an ensemble average of the matrix elements $M_{\gamma\delta}$ given in equation (3.2), and n_0 is the impurity concentration per unit length. These quantities will now be calculated for the model wire considered in the previous chapter.

3.3 The single particle density of states.

The evaluation of this self-energy is still not an easy task. To facilitate the calculation we will therefore make one last approximation and justify it later. The real part of the self-energy which shifts the energy of the excitations is usually small and so will neglect it. The imaginary component, Γ, may then be calculated iteratively from equation (3.12b). For the δ– function scattering potentials, the averaged matrix elements are k independent and so the self-energy is dependent only on the sub-band labels and on the energy. The quantity of interest is the retarded self-energy (which may be obtained from the above through the analytic continuation $ip_n \rightarrow \varepsilon + i\delta$). As an illustration of the type of behaviour arising, we have plotted $\Gamma(\varepsilon=0)$ (the value at the Fermi surface) as a function of Fermi energy for the lowest sub-band in figure (3.1). In order to specify the degree of disorder we have proceeded as in the previous chapter by calculating the parameter $k_F l$ for a Fermi energy of 2.5 meV. Notice that Γ is largest in the region of the sub-band minima and so the broadening will also be largest in this region. As the impurity concentration n_0 decreases, so does Γ, and eventually as n_0 tends to zero, Γ becomes vanishingly small everywhere and we regain the unperturbed form for the Green's function (equation 3.11). The scaling of $\Gamma(0)$ with n_0 is linear if ε_F lies well away from any of the sub-band minima. Near the minima however where Γ is peaked, this scaling is not so trivial, and as the impurity concentration increases the peaks are progressively displaced upwards in energy whilst also becoming broader.

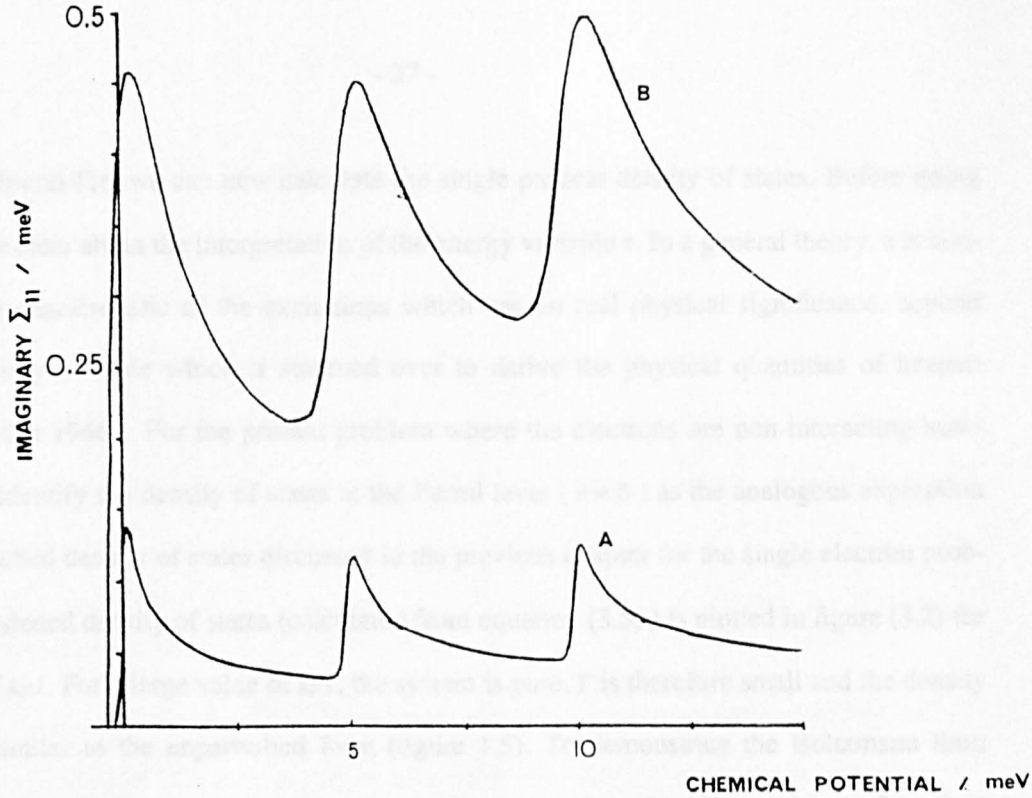


Figure 3.1: Imaginary part of the self-energy in the lowest sub-band as a function of chemical potential for A) $k_F l = 30$ and B) $k_F l = 5$.

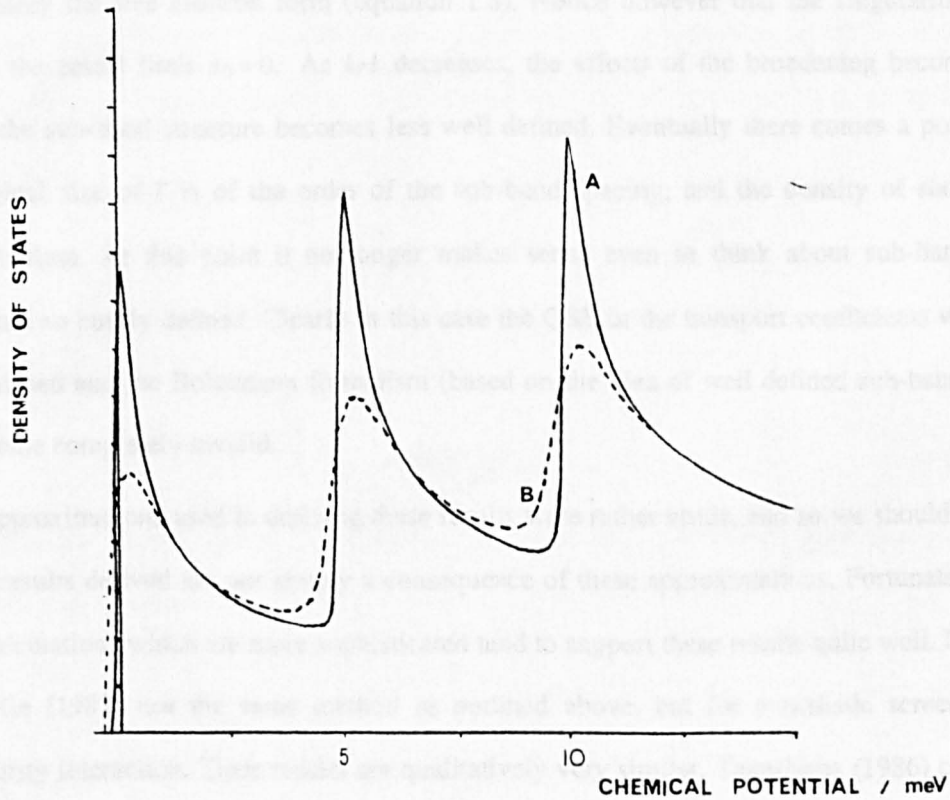


Figure 3.2: The single-particle density of states as a function of chemical potential A) $k_F l = 30$, B) $k_F l = 5$.

Having found $\Gamma(\varepsilon)$ we can now calculate the single particle density of states. Before doing so, we must be clear about the interpretation of the energy variable ε. In a general theory, ε is simply an energy characteristic of the excitations which has no real physical significance, beyond that of a dummy variable which is summed over to derive the physical quantities of interest (Bonch-Bruevich 1966). For the present problem where the electrons are non-interacting however, we can identify the density of states at the Fermi level ($\varepsilon = 0$) as the analogous expression to the unperturbed density of states discussed in the previous chapter for the single electron problem. The broadened density of states (calculated from equation (3.5)) is plotted in figure (3.2) for two values of $k_F l$. For a large value of $k_F l$, the system is pure, Γ is therefore small and the density of states is similar to the unperturbed form (figure 1.5). To demonstrate the Boltzmann limit analytically, we note that $\Gamma \to 0^+$ as $k_F l \to \infty$, so we may replace $\mathrm{Im}\, G^R_{\alpha\alpha}(k,\varepsilon)$ by $-\pi\,\delta(\varepsilon - \xi_{\alpha k})$ whereupon,

$$g(\varepsilon) = \sum_{\alpha k} \delta(\varepsilon - \xi_{\alpha k}) \tag{3.13}$$

$$g(\varepsilon{=}0) = \frac{L}{\pi}\left[\frac{m^*}{2\hbar^2}\right]^{1/2} \sum_{\alpha} \frac{\theta(\varepsilon_F - \varepsilon_\alpha)}{(\varepsilon_F - \varepsilon_\alpha)^{1/2}}$$

which is precisely the free electron form (equation 1.8). Notice however that the singularities only occur in the actual limit $n_0 = 0$. As $k_F l$ decreases, the effects of the broadening become apparent and the sub-band structure becomes less well defined. Eventually there comes a point where the typical size of Γ is of the order of the sub-band spacing, and the density of states becomes featureless. At this point it no longer makes sense even to think about sub-bands because they are so poorly defined. Clearly in this case the QSE in the transport coefficients will also have vanished and the Boltzmann formalism (based on the idea of well defined sub-bands) will have become completely invalid.

Many approximations used in deriving these results were rather crude, and so we should be sure that the results derived are not simply a consequence of these approximations. Fortunately, subsequent calculations which are more sophisticated tend to support these results quite well. Das Sarma and Xie (1987) use the same method as outlined above, but for a realistic screened electron-impurity interaction. Their results are qualitatively very similar. Takeshima (1986) con-

sidered electron-electron correlation effects as well, and took the calculation to much higher order in the impurity interaction. His results are still essentially the same as ours, and would certainly lead to the same type of behaviour in the transport coefficients. One interesting feature of this higher order calculation is that the real part of the self-energy has the effect of displacing the peaks in the density of states upwards in energy, but all by essentially the same amount. Thus it appears that Re Σ does amount to little more than a shift of the origin of energy, and suggests that by ignoring it we have not made too vital an assumption.

3.4 The electrical conductivity.

A closed expression for the electrical conductivity of a disordered solid was first written down by Kubo (1956) and Greenwood (1958), and the first systematic evaluations carried out by Edwards (1958) and subsequently Langer (1960). Our derivation will follow that found in Mahan (1981) but with modifications to allow for the sub-band structure in a wire. The current-current correlation function,

$$\pi_{\gamma\delta}(i\omega_n) = -\frac{1}{L}\int_0^\beta d\tau\, e^{i\omega_n\tau} < T_\tau\, j_\gamma(\tau)\, j_\delta(0) > \qquad (3.14)$$

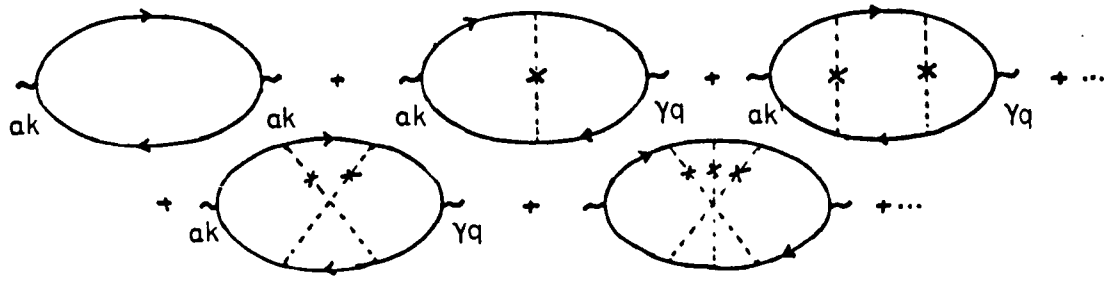
is the two-particle function which describes the linear response of the system to an applied electric field. A suitable form for the spatially averaged current operator when the field acts along the length of the wire is,

$$j_z = -\frac{|e|\hbar}{m^*}\sum_{\alpha,k} k\, C^\dagger_{\alpha k}\, C_{\alpha k} \qquad (3.15)$$

The electrical conductivity is obtained from the retarded correlation function $\pi^R(\omega)$, and the zero frequency limit yields the DC value,

$$\sigma = -\lim_{\omega\to 0}\left[\frac{\mathrm{Im}[\pi^R(\omega)]}{\omega}\right] \qquad (3.16)$$

Expanding equation (3.14) using the S-matrix produces the following diagrams after configuration averaging has taken place,



where the full lines are damped Matsubara functions and the off-diagonal terms are once again taken to be small. The rules for these diagrams are exactly the same as for the self-energy diagrams. These two-particle terms describe the correlation between particle-like (upper line) and hole-like (lower line) excitations propagating in the presence of disorder.

The upper series of diagrams shown in this particle-hole channel (the ladder diagrams) are known to lead to the Boltzmann result in the limit of low disorder (Edwards 1958). For δ – function scattering potentials only the first of these diagrams is non-zero after the summation over momentum is carried out. This simple bubble graph then provides the leading order contribution to the conductivity, and retaining only this diagram will give the broadened equivalent of the Boltzmann conductivity. We therefore write,

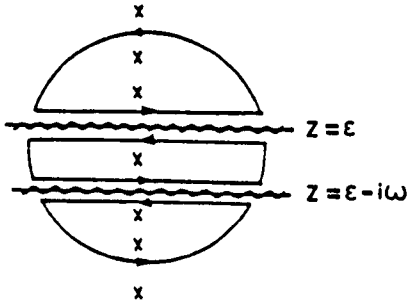
$$\pi^0(i\omega) = \frac{2e^2\hbar^3}{(m^*)^2 L} \sum_{\alpha,k} k^2 \frac{1}{\beta} \sum_{ip_n} G_{\alpha\alpha}(k, ip_n + i\omega_n)\, G_{\alpha\alpha}(k, ip_n) \qquad (3.17)$$

The evaluation of this is not difficult, but it is nevertheless instructive to illustrate the ideas we shall use in later calculations (detailed discussions are given in Mahan 1981). The first step is to perform the summation over the Matsubara frequencies $ip_n = (2n+1)\pi i/\beta$, which is achieved by considering the contour integral,

$$I = \int \frac{dZ}{2\pi i} f(Z)\, G(Z)\, G(Z + i\omega_n) \qquad (3.18)$$

The contour will be defined shortly. (We have temporarily suppressed the k and α labels; they will be reinserted at the end). The function $f(Z)$ is simply the Fermi function: $f(Z) = (e^{\beta Z} + 1)^{-1}$, and since this has poles at $Z = ip_n$, the residues of equation (3.18) will be the terms in our summation. An evaluation of this integral will thus give us the sum over the Matsubara frequencies. The self-energy and hence Green's functions are not analytic on the real axis, so there are branch cuts at $Z = \varepsilon$ and $Z = \varepsilon - i\omega_n$ (with epsilon real). An important point is that the frequencies $i\omega_n$ which

refer to the two particle function are Bose (even) frequencies and so do not lie on the $i\omega_n$ poles. The contour for the integral is therefore,



to be evaluated in the limit $R \to \infty$. By Jordan's lemma, the contribution from the arc vanishes in this limit and we are left with,

$$I = \int \frac{d\varepsilon}{2\pi i} f(\varepsilon)\ [G(\varepsilon + i\omega_n)(G(\varepsilon + i\delta) - G(\varepsilon - i\delta)) + G(\varepsilon - i\omega_n)(G(\varepsilon + i\delta) - G(\varepsilon - i\delta))] \tag{3.19}$$

Noting that,

$$G(\varepsilon + i\delta) - G(\varepsilon - i\delta) = G^R(\varepsilon) - G^A(\varepsilon) = -iA(\varepsilon)$$

where $A(\varepsilon)$ is the spectral function, gives us for for the correlation function

$$\pi_0(i\omega_n) = \frac{2e^2\hbar^3}{(m^*)^2 L} \sum_{\alpha,k} k^2 \int \frac{d\varepsilon}{2\pi} f(\varepsilon) A(\varepsilon)(G(\varepsilon + i\omega_n) - G(\varepsilon - i\omega_n)) \tag{3.20}$$

Next we take the analytic continuation $i\omega_n \to \omega + i\delta$ to generate the retarded function, and take the limit $\omega \to 0$ to find the DC conductivity. Carrying out these steps and noting,

$$\lim_{\omega \to 0} \frac{1}{\omega}(f(\varepsilon) - f(\varepsilon+\omega)) = -\frac{\partial f}{\partial \varepsilon}$$

gives us the final expression for σ,

$$\sigma = \frac{e^2\hbar^3}{m^{*2}} \sum_{\alpha} \int \frac{k^2 dk}{2\pi} \int \frac{d\varepsilon}{2\pi} A_\alpha(k,\varepsilon)^2 \left[-\frac{\partial f}{\partial \varepsilon} \right] \tag{3.21}$$

This equation represents the broadened version of the Boltzmann conductivity, and explicitly contains the spectral function of the system. To demonstrate the Boltzmann limit we again assume the limit $\Gamma \to 0$ and note that to lowest order,

$$\lim_{\Gamma \to 0} A^2 = \lim_{\Gamma \to 0} \left[\frac{2\Gamma}{x^2 + \Gamma^2} \right]^2 = \frac{2\pi\delta(x)}{\Gamma(\varepsilon)}$$

where $x = \varepsilon + \mu - \varepsilon_\alpha - \hbar^2 k^2/2m^*$. If we associate in the usual fashion $\Gamma = \hbar/2\tau$, equation (3.21) reduces to (at zero temperature),

$$\sigma = \frac{2e^2}{\pi}\left[\frac{2m^*}{\hbar^2}\right]^{1/2}\sum_{\alpha}\frac{(\varepsilon_F-\varepsilon_\alpha)^{1/2}\,\theta(\varepsilon_F-\varepsilon_\alpha)\,\tau_\alpha(\varepsilon_F)}{m^*}$$

which is identical to the Boltzmann result (equation (2.16)). Both expressions are therefore the same at finite temperatures as well, since both involve only an integral over $\partial f/\partial\varepsilon$ to obtain the finite temperature result. Let us be clear upon the meaning of the timescale we have introduced through Γ. This timescale $\tau = \hbar/2\Gamma$ relates to the damping of a quasi-particle in the presence of scattering, and characterises the time over which it is meaningful to label an excitation as having momentum k and being in sub-band α. The timescale entering the conductivity on the other hand refers to the two-particle function and is a transport time. In general these two timescales are not the same as each other, and indeed, may be very different (Das Sarma and Stern 1985). They are only the same here accidently, because the scatterers are assumed to have zero range and the scattering is elastic. Had the potential range been finite, the scattering would have been k dependent whereupon the ladder diagrams in the particle-hole channel would have been non-zero. The timescale entering the Boltzmann limit would then no longer have been $\hbar/2\Gamma$ but the usual transport time (Butcher 1973, 1986).

Before solving equation (3.21) numerically, it is worthwhile briefly considering its analytical structure. The factor relating to the spectral function may be written in the form,

$$A^2(k,\varepsilon) = 2\,G^R(k,\varepsilon)\,G^A(k,\varepsilon) - 2\mathrm{Re}[(G^R(k,\varepsilon))^2]. \qquad (3.22)$$

This is exactly the form that the results of the zero temperature calculations of Langer (1960) and Cantrell and Butcher (1985) are written in. When the impurity concentration is very low, the first term in this expression dominates and the second may be neglected (the presence of the second term does ensure though that all the singularities are cut-off in the k integrals). This is a standard approximation made when evaluating the conductivity, however if we explicitly want to discuss broadening the second term should be retained. Later on though we shall perform some calculations ignoring this second term. Within the framework of the present model it is possible to perform the sum over k by a contour integral. The answer however is complicated and no more intuitive than the above expressions for the conductivity, which we therefore choose to evaluate numerically.

A plot of the fully broadened conductivity as a function of chemical potential is shown in figure (3.3). We have dispensed with the parameter $k_F l$ (which is a little abstract) and instead, have specified the degree of disorder by calculating the mobility of the device ($\sigma = ne\mu$). This makes comparison to experiment a little easier. The simplest experiments to perform in which the chemical potential may be varied are in silicon MOSFETS (Hartstein et al 1984, Skocpol et al 1984, Wheeler et al 1984, Warren et at 1986), and so the results we calculate are pertinent to these devices. The effective mass used is $0.2m_0$ (Ando et al 1982), and we quote the mobility at a Fermi energy of 12.5 meV, where the areal density of electrons is $\sim 2\times10^{12} cm^{-2}$ and is typical of these devices. The two mobilities chosen are 3000 cm^2/Vs (which is quite low) and 20,000 cm^2/Vs which is rather high. In figure (3.3) we see that in the high mobility device QSE effects are still predicted, though even here they are much smaller than Boltzmann theory predicts. In the low mobility system however the QSE are almost entirely absent because of the lifetime broadening. In the light of this it is hardly surprising that the experimental results discussed previously show so little structure. In addition, we have chosen devices with sub-bands that are spaced by 5 meV (corresponding to a device width of $\approx 20nm$). Wider channels mean correspondingly closer spaced sub-bands; for example the Warren device (1986) has a width of order $50nm$ and a sub-band separation of about 2 meV. In this case the effect of the broadening is enhanced since the degree of mixing between the sub-bands is larger, and so the QSE are correspondingly reduced. Our calculations show that the structure found by Skocpol et al (1984) and Warren et al (1986) is consistent with the effect of broadening due to the impurities. Subsequent calculations by Das Sarma and Xie (1987) support these conclusions for screened electron-impurity scattering rather than phenomenological δ – function potentials. The conclusion we draw from these results is that devices currently available are just on the borderline of being able to resolve sub-band structure in the conductivity, insofar as localisation and fluctuations can be ignored. After we have discussed these effects we shall be able to quantify this statement further.

3.5 The Chester-Thellung theorem.

The calculation of thermal transport coefficients via the Kubo formula is in general rather

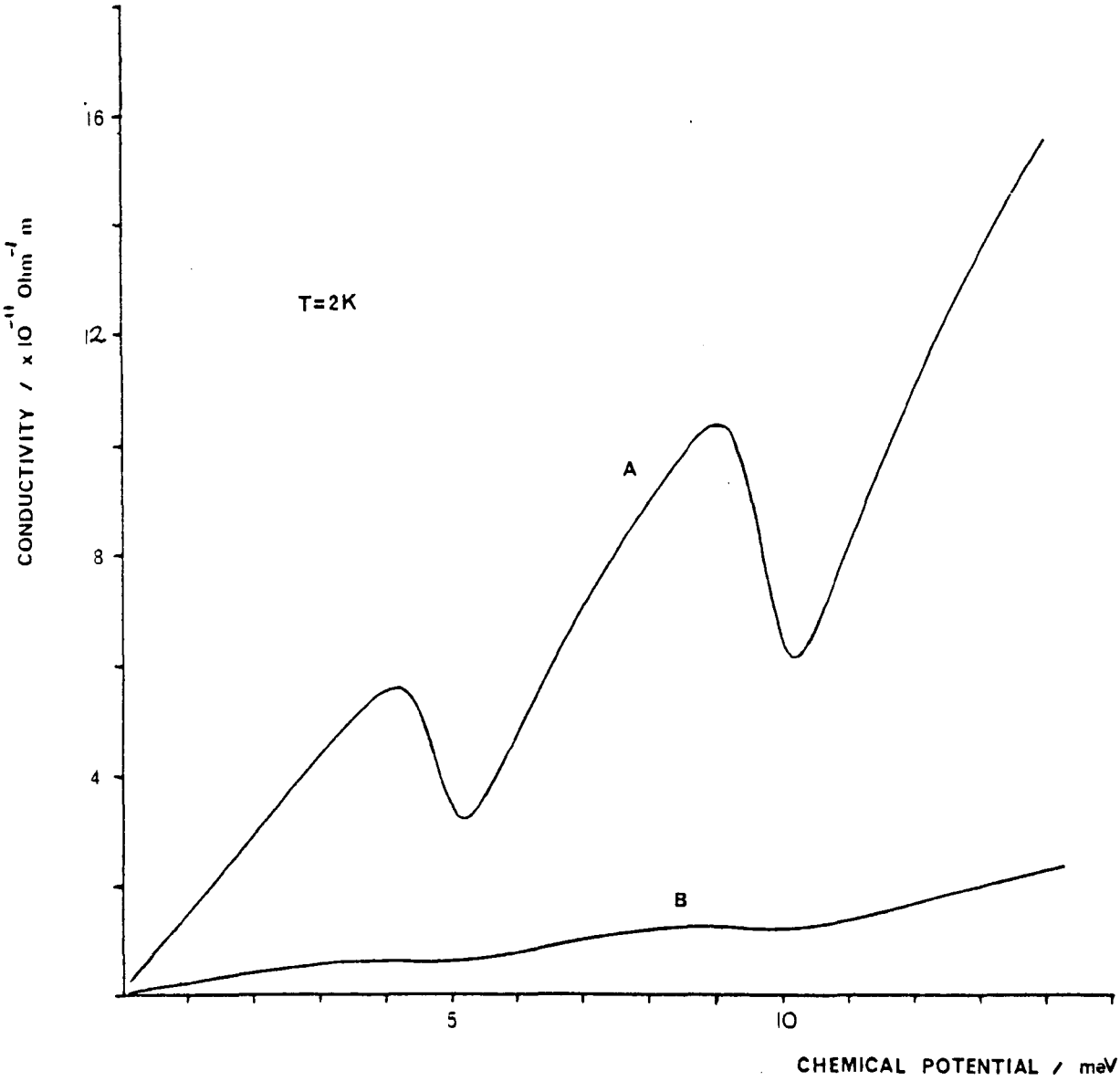


Figure 3.3: The electrical conductivity against chemical potential for two different mobilities A) $20000 cm^2/Vs$ and B) $3000 cm^2/Vs$.

difficult (Mahan 1981, Vilenkin and Taylor 1978, Stedman and Kaiser 1987) owing to the complexity of deriving the correct heat current operator. Chester and Thellung (1961) however were able to derive an exact relationship between the correlation functions governing thermal transport, within certain restrictions. Starting from the Kubo formula, their proof rests upon assuming that the electrons are non-interacting and that the scattering is elastic, but is otherwise quite general. In particular, it does not depend upon any assumption about the strength of the disorder. In the context of our present problem it is therefore very useful because it enables us to express (exactly) the thermal transport coefficients in terms of the zero temperature conductivity. Although we shall consider only the the isotropic case again, Smrcka and Streda (1977) have extended the proof to show its validity for anisotropic systems as well.

We will sketch only an outline of the proof. The coefficients $L_{\alpha\beta}$ defined in equations (2.1a) and (2.1b) are given by the general Kubo formula,

$$L_{\alpha\beta} = -\int_0^\infty dt\ e^{-st} \int_0^\infty d\beta'\ Tr[\rho_0 j_\beta(-t-i\beta')\, j_\alpha(0)] \qquad s \to 0^+ \tag{3.23}$$

where ρ_0 is the equilibrium density matrix. The operator j_1 is the electric current operator and j_2 is the heat current operator. (As defined here, the $L_{\alpha\beta}$ are slightly different to the S_{ij} used by Chester and Thellung, though each set leads to the same expressions for the physical observables). Of crucial importance to the argument is that the total Hamiltonian is separable into a sum of single particle Hamiltonians, since the electrons are assumed non-interacting. If we define the complete set of one-electron eigenfunctions $|n>$ such that,

$$H\,|n> = E_n\,|n>$$

then it is clear that the separability of the total Hamiltonian implies that ρ_0 is diagonal in this representation. This leads to significant simplifications in the formal manipulation of equation (3.23). Using the $|n>$ as the basis of the representation, the operators may be cast into second quantised form,

$$j_\alpha(0) = \sum_{nn'} <n|j_\alpha|n'> C_n^\dagger C_{n'} \tag{3.24a}$$

$$j_\beta(-t-i\beta') = \sum_{mm'} <m|j_\beta|m'> e^{-(it-\beta')(E_m - E_{m'})} C_m^\dagger C_{m'} \tag{3.24b}$$

The integrals over t and β' are trivial, as is the evaluation of the trace since,

$$Tr[\rho_0 C_m^+ C_{m'} C_n^+ C_{n'}] = f_m(1-f_n)\,\delta_{n'm}\,\delta_{m'n} \tag{3.25}$$

$$f_m = Tr[\rho_0 C_m^+ C_m]$$

where f_m is the usual Fermi-Dirac distribution function. The term with $m=m', n=n'$ is not considered; this is a subtlety necessary to ensure that the transport coefficients are finite, and does not result in any serious loss of generality (Greenwood 1958, Chester and Thellung 1961). Combining equations (3.23), (3.24) and (3.25) gives,

$$L_{\alpha\beta} = -\pi \sum_{m,n} \frac{\partial f}{\partial E_n} <m|j_\beta|n><n|j_\alpha|m> \delta(E_m - E_n) \tag{3.26}$$

As defined here, the electric current operator $j_1 = -|e|j$ (where j is the particle current operator) and the heat current operator $j_2 = (j_E - \mu j)$ where the energy current,

$$j_E = \frac{1}{2}(Hj + jH)$$

Using this result, and letting the size of the system $L \to \infty$ so that the energy levels become quasi-continuous and the summation may be replaced by an integral, gives,

$$L_{11} = \int \sigma(\varepsilon)\,(-\partial f/\partial\varepsilon)\,d\varepsilon \tag{3.27a}$$

$$L_{12} = L_{21} = -\frac{1}{|e|}\int \sigma(\varepsilon)\,(\varepsilon-\mu)\,(-\partial f/\partial\varepsilon)\,d\varepsilon \tag{3.27b}$$

$$L_{22} = \frac{1}{|e|^2}\int \sigma(\varepsilon)\,(\varepsilon-\mu)^2\,(-\partial f/\partial\varepsilon)\,d\varepsilon \tag{3.27c}$$

and,

$$\sigma(\varepsilon) = \pi\,\rho(\varepsilon) \sum_m |<n|j_1|m>|^2\,\delta(\varepsilon - E_m)$$

where $\rho(\varepsilon)$ is the density of states. This is the Chester-Thellung result. Notice that at no point has it been necessary to assume anything about the strength of the disorder. As before, the function $\sigma(\varepsilon_F)$ is just the zero temperature conductivity of the system, and in general this will be a very complicated function of its arguments. The important point however is that the function $\sigma(\varepsilon)$ occurs in the integrand of all the $L_{\alpha\beta}$. Compare this to the relaxation time result (equations 2.15 and 2.16). The result is exactly the same, but we have now established a much wider range of validity. The very abstract nature of the proof leads us to suspect that the above results are general to all disorders and dimensionalities, and we shall make extensive use of this hypothesis later.

3.6 Thermal conduction and thermopower in disordered wires.

Using the Chester-Thellung relationships we can now calculate the thermal transport coefficients in the devices considered above. The effect of broadening on the QSE in the conductivity was particularly severe. This is also true of the thermal conductivity (figure 3.4), where the size effects are even less observable, even in the high mobility device. Given the experimental difficulty of measuring κ in very small systems, it seems unlikely that it will prove possible to observe QSE in the thermal conductivity. The thermopower on the other hand is much less sensitive to the effects of the broadening (Cantrell and Butcher 1985c). In the high mobility ($20{,}000 cm^2/Vs$) device (figure 3.5), the structure is almost the same as that predicted by Boltzmann theory. As the temperature is raised, so the positive peaks become smaller, but even at $8K$ considerable structure is visible. In the lower mobility ($3000 cm^2/Vs$) system (figure 3.6) the evidence of the sub-band structure is still apparent, despite the fact that for these parameters the electrical and thermal conductivities are almost featureless. These results are relevant to the Warren device (1986), and we predict that in many of the quasi-1D MOSFET devices so far discussed quantum size effects will be observable in the thermopower. Figure (3.7) shows the variation of the thermopower with temperature for the two devices and two different Fermi levels. Again we have allowed for the variation of the chemical potential with temperature by calculating the number of electrons in the system (using the perturbed density of states) and ensuring that it remains constant. The conclusion is that it should be possible to observe a positive contribution to the diffusive thermopower for temperatures below about $6K$, if the Fermi energy lies in the vicinity of a sub-band minimum and the sub-bands are relatively well separated, even if the electrical conductivity shows no evidence for QSE at all.

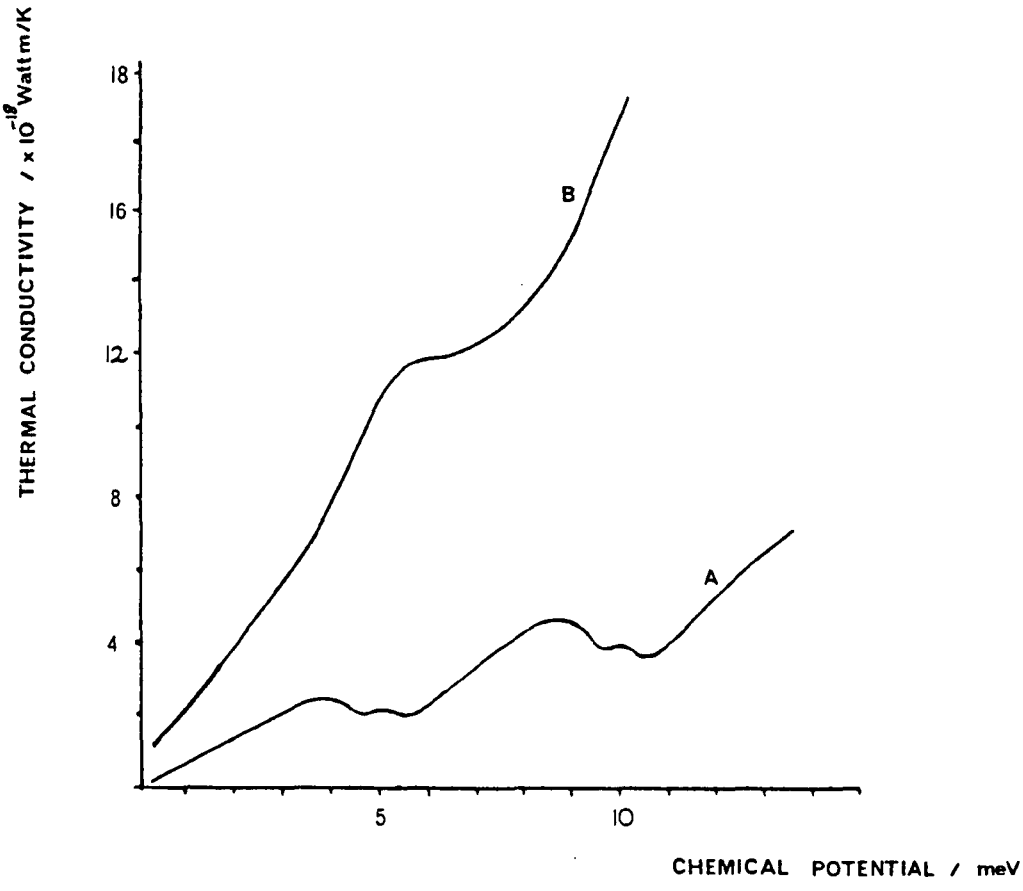


Figure 3.4a: Thermal conductivity versus chemical potential for a mobility of $20000 cm^2/Vs$ and temperatures of A) $T = 2K$ and B) $T = 8K$.

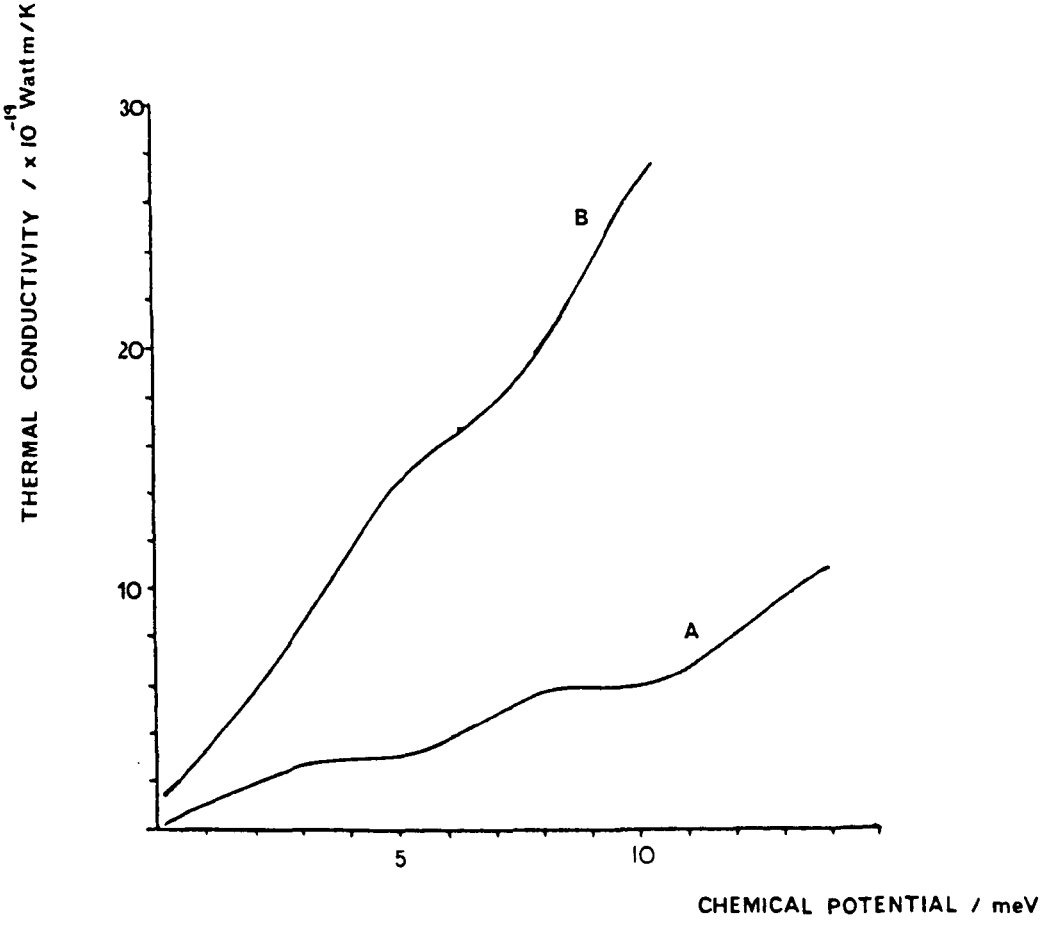
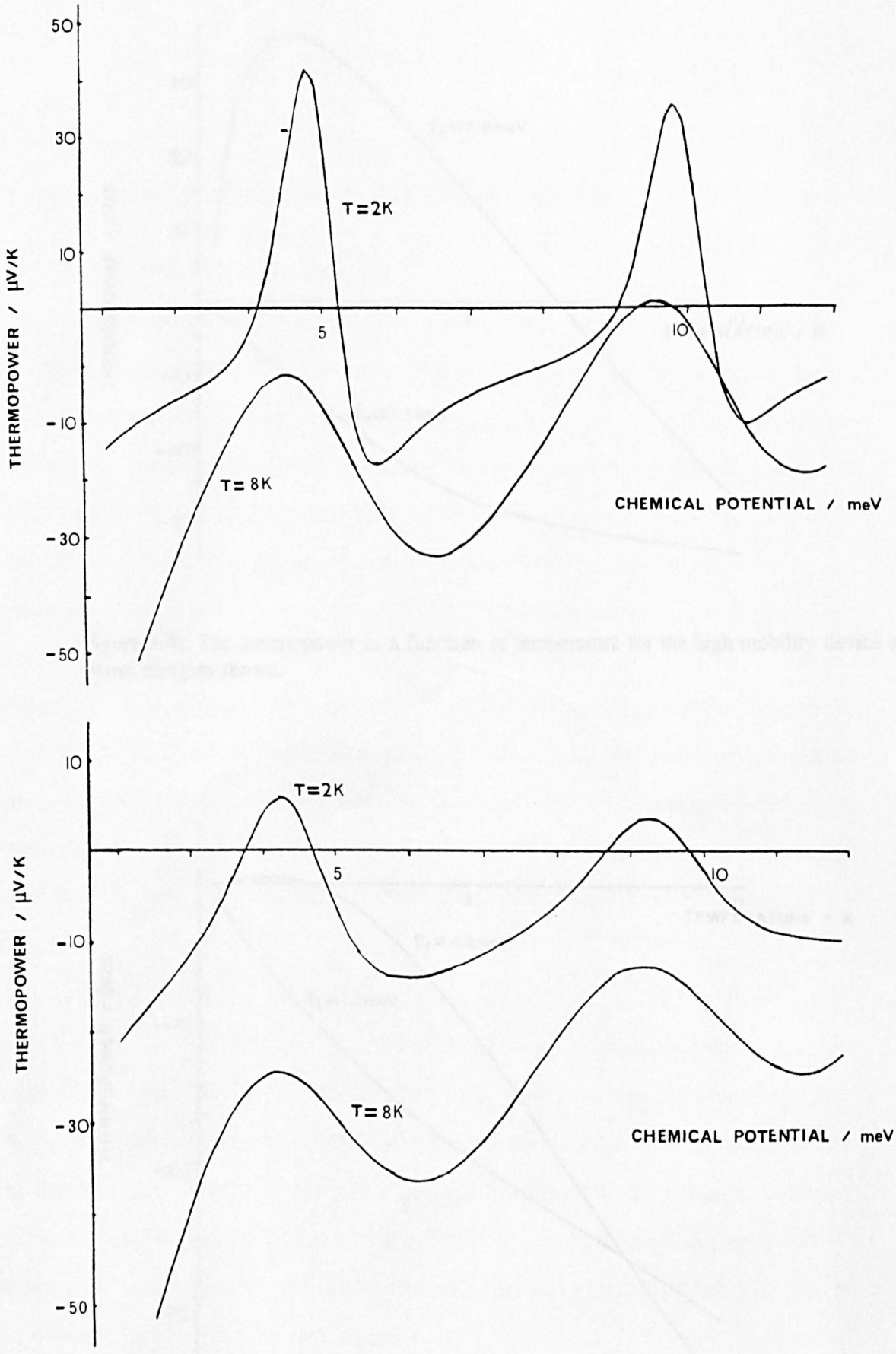


Figure 3.4b: Thermal conductivity versus chemical potential for a mobility of $3000 cm^2/Vs$ and temperatures of A) $T = 2K$ and B) $T = 8K$.



Figures 3.5 and 3.6: The thermopower as a function of chemical potential for the temperatures indicated. The upper graph is the high mobility case ($20000 cm^2/Vs$) and the lower corresponds to the low mobility case ($3000 cm^2/Vs$).

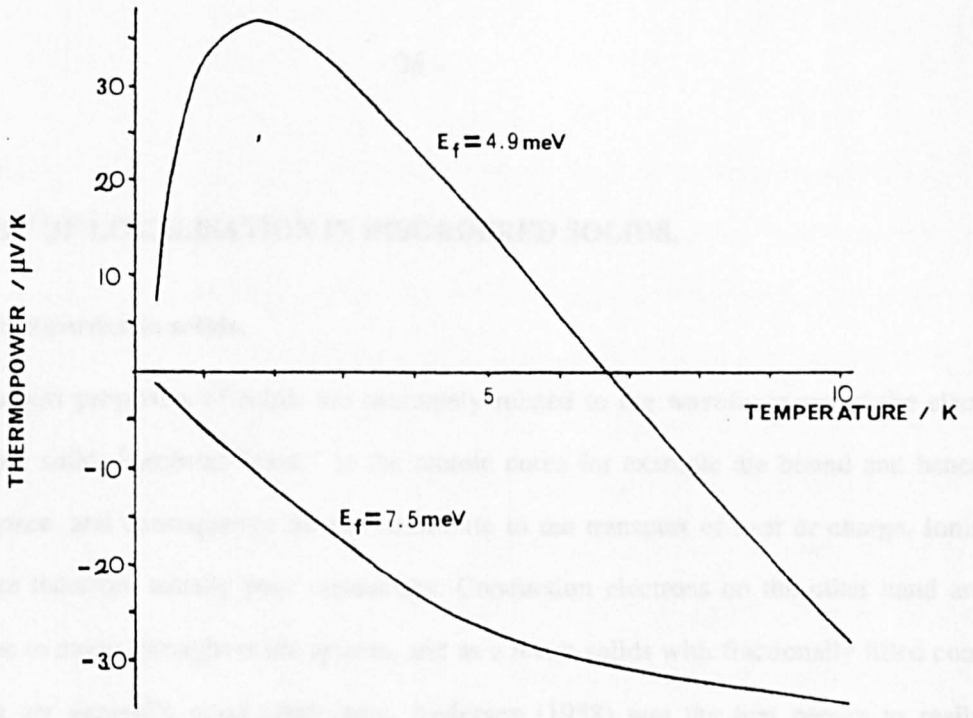


Figure 3.7a: The thermopower as a function of temperature for the high mobility device and Fermi energies shown.

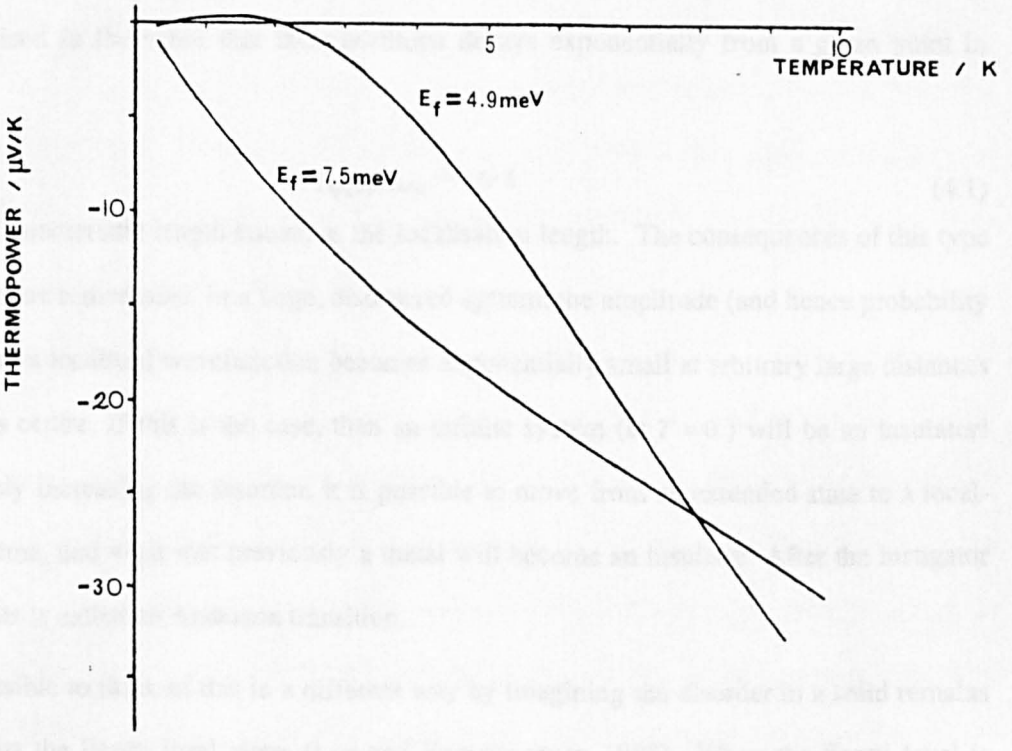


Figure 3.7b: The thermopower as a function of temperature for the low mobility device and Fermi energies shown.

CHAPTER 4

THE THEORY OF LOCALISATION IN DISORDERED SOLIDS.

4.1 The role of disorder in solids.

The transport properties of solids are intimately related to the wavefunctions of the electrons within the solid. Electrons "close" to the atomic cores for example are bound and hence localised in space, and consequently do not contribute to the transport of heat or charge. Ionic compounds are therefore usually poor conductors. Conduction electrons on the other hand are essentially free to move throughout the system, and as a result solids with fractionally filled conduction bands are generally good conductors. Anderson (1958) was the first person to really address the question as to how wavefunctions behave in a disordered medium. The simplest idea is that disorder merely results in Bloch-like (if we are discussing the conduction electrons) wavefunctions losing coherence due to scattering, but otherwise remaining extended throughout the system. As a result, a metal will conduct however strong the disorder, though the magnitude of the conductivity will decrease as the scattering increases. Anderson showed however that this simple picture is not correct; if the disorder is sufficiently strong electronic wavefunctions may become localised in the sense that their envelope decays exponentially from a given point in space,

$$|\psi(\mathbf{r})| \sim e^{-|\mathbf{r}-\mathbf{r}_0|/\xi} \tag{4.1}$$

where ξ is a characteristic length known as the localisation length. The consequences of this type of localisation are remarkable. In a large, disordered system, the amplitude (and hence probability of diffusing) of a localised wavefunction becomes exponentially small at arbitrary large distances away from its centre. If this is the case, then an infinite system (at $T = 0$) will be an insulator! Thus by simply increasing the disorder, it is possible to move from an extended state to a localised state regime, and what was previously a metal will become an insulator. After the instigator of the idea, this is called an Anderson transition.

It is possible to think of this in a different way by imagining the disorder in a solid remains constant whilst the Fermi level alters (Lee and Ramakrishnan 1985). When the Fermi level is

low the electrons will "see" the potential fluctuations due to the disorder and may become localised. When the Fermi level is high on the other hand the electrons see less of the disorder potential and so their chance of remaining extended is increased. As the Fermi level (or correspondingly the electron concentration) is altered therefore, it is possible to imagine moving from an extended (metallic) to a localised (insulator) regime, which is known as the Metal-Insulator transition. The energy which marks the boundary between localised and extended states is believed to be sharp and is called the mobility edge ε_c. This energy obviously marks the boundary between conducting and insulating regimes. A natural question to ask then is exactly how does the conductivity go to zero as the Fermi energy approaches ε_F from the metallic side? Assuming that the Boltzmann formula (equation 1.1) remains valid for ε_F arbitrarily close to ε_c on the metallic side, and that the lowest sensible limit for the mean free path is the interatomic separation a, suggests that the transition is discontinuous and there that is a minimum metallic conductivity of the form,

$$\sigma_{\min}^{3D} \sim \frac{1}{3\pi^2} \frac{e^2}{\hbar} \frac{1}{a} \tag{4.2}$$

There is a great deal of experimental evidence for the existence of a Metal-Insulator transition in disordered 3D systems (see the articles by Mott et al (1975), Mott and Davies (1979), Ando et al (1982) and Milligan et al (1985)). On the other hand, careful experiments have questioned the existence of σ_{min}, with conductivities lower than σ_{min} having been measured (see the beautiful experiments on phosphorus doped silicon by Paalanen et al 1982, 1983). Experiments such as these argue the case that the transition is more likely to be continuous than discontinuous, even though it is very sharp near the mobility edge.

Although the concept of a minimum metallic conductivity is easy to understand, there are disturbing theoretical (as well as experimental) reasons for believing it is in fact incorrect. Firstly it assumes, without justification, that the Boltzmann formula is universally valid on the metallic side of the transition. Secondly it assumes that localisation effects are entirely absent on the metallic side, but are present on the insulating side leading to a divergent localisation length as $\varepsilon \rightarrow \varepsilon_c$. Abrahams et al (1979) proposed a scaling theory of the Metal-Insulator transition, describing it within the language of phase transitions and critical phenomena, which neatly

sidestepped these problems. Near the mobility edge they argued that the only relevant length scales are a correlation length on the metallic side and the localisation length on the insulating side, both of which diverge as $\varepsilon \rightarrow \varepsilon_c$ from above and below. The conclusion which is drawn from this is that the conductivity must go to zero continuously at the mobility edge, and as a result there is no σ_{min}.

The scaling theory of localisation (which we shall look at shortly) highlights the importance of the various length scales in the problem, particularly in relation to the localisation length ξ. If the system size L is very large ($L \gg \xi$), then the likelyhood of diffusion across the sample vanishes exponentially with L. This is known as the strongly localised regime. In the opposite limit ($L \ll \xi$), the wavefunction is to all intents and purposes extended throughout the sample. We anticipate that in this weakly localised regime, the conductivity will be well described by a Boltzmann like term plus a (size dependent) correction due to the localisation. As the system size increases, the size dependent corrections will either decrease so that as $L \rightarrow \infty$ we recover the metallic limit, or increase until the conductivity saturates to zero, depending upon the disorder in the system. This implies the existence of a critical disorder, below which $\sigma \rightarrow \sigma_B$ as $L \rightarrow \infty$ and above which $\sigma \rightarrow 0$ as $L \rightarrow \infty$. This critical disorder may be specified by the conductance at a length scale of the mean free path l.

4.2 Consequences of low dimensionality.

The arguments so far have been applicable to disordered three dimensional solids. Very quickly it also became clear that dimensionality is a crucial variable in determining the extent of the localisation. For example, Mott and Twose (1961) were able to show quite generally that in one dimension all the states are localised no matter how weak the disorder, due to repeated backscattering. In the limit of weak disorder the localisation length in 1D is therefore of the order of the mean free path (Thouless 1973). As a result, all infinite one-dimensional disordered systems are insulators at T=0, except for very unlikely (i.e. periodic) impurity configurations. Naturally then there is no mobility edge and no Metal-Insulator transition in one dimension. The same conclusions have also been reached from different standpoints by Abrikosov and Rhyzkin (1978) and

Landauer (1970). Thouless (1977) considered the case of a wire of finite cross section, and concluded that if its resistance $R > 2\hbar/e^2$ then the states within it would become exponentially localised. The arguments he used were a precursor to the development of the scaling theory, especially in recognising that the nature of the eigenstates could be effectively described by the systems resistance (or equivalently, conductance). For a resistance greater than about $10k\Omega$ therefore, it is impossible to have metallic conduction at $T = 0$, and the resistance of a wire will grow exponentially rather than linearly with its length.

The situation in two dimensions was for a long time rather unclear (scaling theory shows that d=2 is a critical dimension for localisation). Licciardello and Thouless (1975a,b) extended Mott's σ_{min} idea to two dimensions to predict,

$$\sigma_{min}^{2D} \sim 0.1 e^2/\hbar \tag{4.3}$$

Simply on dimensional grounds there is no length scale which enters σ_{min}^{2D}, which is therefore a universal quantity for all materials. This is the first sign that d=2 is in some sense special. Their numerical simulations on a range of differing types of $2D$ lattice supported the existence of this universal minimum conductivity rather well. The situation experimentally was reviewed in detail by Pepper (1977), who showed that this simple universality was not quite correct however. Again we should worry about the fact that there is no a priori reason for assuming there is a region of extended states in $2D$. We could equally well have assumed the same in 1D where it is known that all the states are localised. The nice feature of the scaling theory is that it allows the dimensionality to be treated as a continuous variable and so questions such as these can be tackled. In 2D the prediction is that, in fact, all the states are localised, although the localisation length can be very large in pure samples and diverges exponentially as the degree of disorder goes to zero. This explains why the numerical simulations were misinterpreted, because the system sizes considered were not really large enough.

4.3 The Scaling theory of localisation.

The ideas behind the scaling theory were developed by Thouless and others (see the references already cited). They asked the question as to what controls the behaviour of a disordered

solid as it doubles its size? The crucial assumption of the scaling theory (Abrahams et al 1979) is that only one parameter is necessary to do this, the conductance of the system at a length scale L. This is not immediately obvious; it could conceivably depend in addition upon the length L itself, the degree of disorder or even the type of material. In the limits which are known from perturbation theory however, the postulate is found to be correct (notwithstanding a challenge to this by Efetov 1985). Following this idea, they constructed a scaling function from the renormalisation group differential recursion relation,

$$\beta(g) = \frac{d \ln g}{d \ln L} \tag{4.4}$$

which is solely a function of the (non-dimensional) conductance g. The procedure is to then use the known limiting forms for the function $g(L)$ to construct the limits of $\beta(g)$, and to extrapolate between them. We have already discussed these two limits. On small length scales, the conductance will be well described in terms of a scale independent Boltzmann conductivity,

$$g(L) = \sigma L^{d-2} \tag{4.5}$$

where d is the system dimensionality. At very large length scales in the insulating regime where $L \gg \xi$ we know that exponentially strong localisation sets in so,

$$g(L) \sim e^{-L/\xi} \tag{4.6}$$

Alternatively we can think about the conductance at large length scales and say that if g is large, then Ohm's law is roughly valid (equation 4.5) and,

$$\beta(g) = d-2 \tag{4.7}$$

If g is very small on the other hand, equation (4.6) is appropriate whereupon,

$$\beta(g) = \ln(g/g_c) \tag{4.8}$$

where g_c is a characteristic conductance of order π^{-2}. Notice that equation (4.8) predicts that the scaling function is negative in this regime, and so as L increases, the conductance rapidly tends to zero. The significance of d=2 is clearly seen in equation (4.7) where $\beta(g)$=0. In order to specify the behaviour here it is necessary to go beyond lowest order perturbation theory and calculate the next term in $\beta(g)$. The result is that (Gor'kov et al 1979, Abrahams et al 1980),

$$\beta(g) = d - 2 - a/g \tag{4.9}$$

where a is a positive quantity for normal isotropic impurity scattering, and is of the order π^{-2}.

The scaling function is now constructed from these two limits by making the plausible assumptions that $\beta(g)$ is both continuous and monotonic. The results are displayed graphically in figure (4.1).

The three dimensional curve contains a point where $\beta(g_3)$ is zero. This an unstable fixed point from which the scaling trajectories flow outwards. Thus if g is slightly greater than g_3, $\beta>0$ and so g increases with L, eventually reaching its Boltzmann value. Conversely if $g<g_3$, then $\beta<0$ and so g decreases with increasing L, the system eventually becomes an insulator. The point $\beta(g_3)=0$ then clearly defines the mobility edge and marks the Metal-Insulator transition. The conductance g_3 marks the minimum metallic conductance (not conductivity) a system can have. The assumption of continuity in the scaling function, β, implies that there is a continuous transition as $\varepsilon \rightarrow \varepsilon_c$ and therefore that there is no minimum metallic conductivity. Furthermore, in the vicinity of the transition we can write on the metallic side (Lee and Ramakrishnan 1985),

$$\sigma \sim \frac{e^2}{\hbar} \frac{1}{\xi_c}$$

where the correlation length $\xi_c \sim (\varepsilon - \varepsilon_c)^{-\nu}$ and ν is a critical exponent which is universal. On the localised side of the mobility edge we can also show that the localisation length ξ_{loc} diverges exactly as ξ_c with the same critical exponent. This has been verified (Paalanen 1983). Furthermore, this is not restricted (in an abstract sense) to $d=3$. The scaling function will have a zero for all $d>2$, and so there will be a Metal-insulator transition for $d>2$. If we write $d = 2+\varepsilon$ and assume $\varepsilon \ll 1$, then we can calculate the perturbative estimate of ν, which is unity (Lee and Ramakrishnan 1985). Obviously this will not be exact for $d = 3$ ($\varepsilon = 1$), and values of order 0.5 are found experimentally (Paalanen 1983).

Two dimensions marks the critical value of d for which $\beta(g)$ has no zero. Scaling theory therefore predicts that there are no extended states in $2D$ and hence no Metal-Insulator transition. As the system size increases, eventually the conductance will tend to zero and the sample will become an insulator. For pure systems this only becomes apparent at truly large length scales, since when g is large, $\beta(g)$ is very small. For many $2D$ systems of interest we can therefore expect $L \ll \xi_{loc}$, whereupon traditional transport theory will provide reasonable answers. The

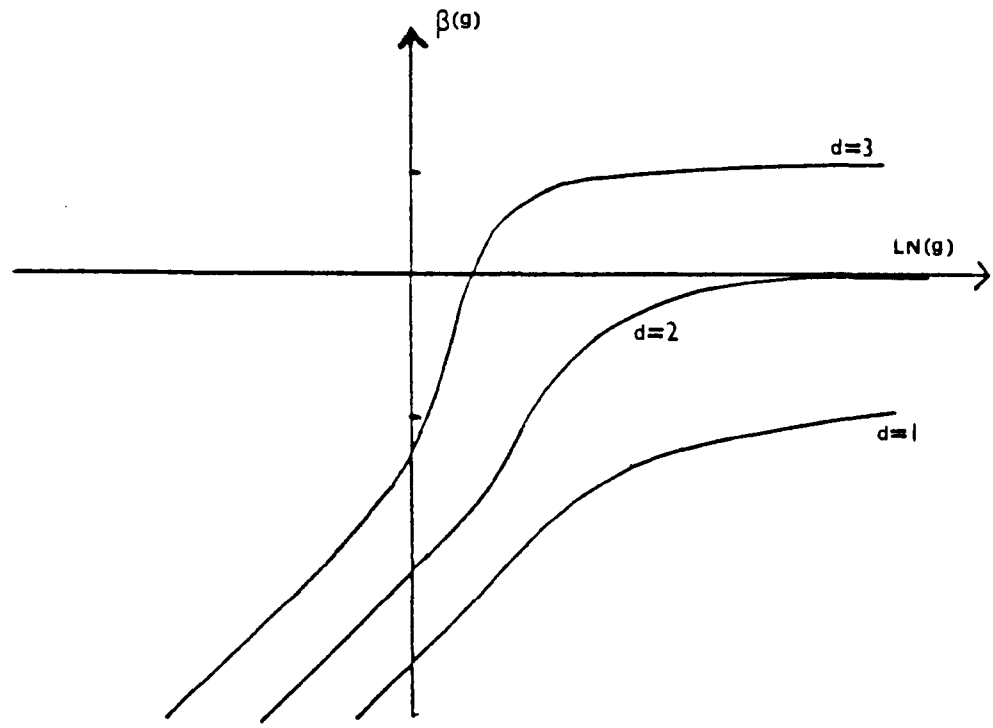


Figure 4.1: A plot of the scaling function $\beta(g)$ against $\ln(g)$ for $d = 1, 2, 3$. Note that β is only positive for large g if $d > 2$, and so there are no extended states for $d \leqslant 2$ (after Abrahams et al 1979).

prediction for one dimension however is that as L increases the conductance rapidly tends to zero. This case will be further discussed in section (4.6). For now we note that scaling theory certainly supports the results of Mott and Twose (1961) and Landauer (1970) that there are no extended states in pure $1D$ and no metallic regime.

4.4 The weak localisation regime.

The scaling theory of localisation gives us a prescription for the behaviour of the conductance at zero temperature of a disordered solid, and leads to concrete predictions such as the existence of certain universal exponents near the mobility edge. In the weakly localised regime we can say a great deal more because the traditional methods of diagrammatic perturbation theory are still valid. In the strongly localised regime, transport at finite temperature occurs via hopping which is calculated in a different manner, by using rate equations, percolation theory and the theory of random walks. For reviews of this regime see Butcher (1976), Shklovskii and Efros (1984). We should also mention in passing that attempts have been made to treat the transition (from the weak to strong localisation) within one theory, using self-consistent perturbation methods (Vollhardt and Wolfle 1980) It is not clear however how to reconcile all of the difficulties which arise when taking this approach.

A helpful picture to consider when thinking about the weakly localised regime ($k_F l \gg 1$) is to think of the electrons executing random walks from impurity to impurity with a mean step l (Al'tshuler and Aronov 1985). Imagine an ensemble of impurities in real space and ask what is the probability for an electron to diffuse in this manner from a point A to another point B (see figure 4.2)? If we sum the amplitudes for all possible paths, then the probability is simply the square modulus of this quantity,

$$Prob. = |\sum_s A_s|^2 \tag{4.10}$$

which may be written as,

$$Prob. = \sum_s |A_s|^2 + \sum_{s \neq t} A_s A_t^* \tag{4.11}$$

The first term is simply the classical sum of all the probabilities to travel along any path. The second term represents an interference of all the various amplitudes which when averaged over a

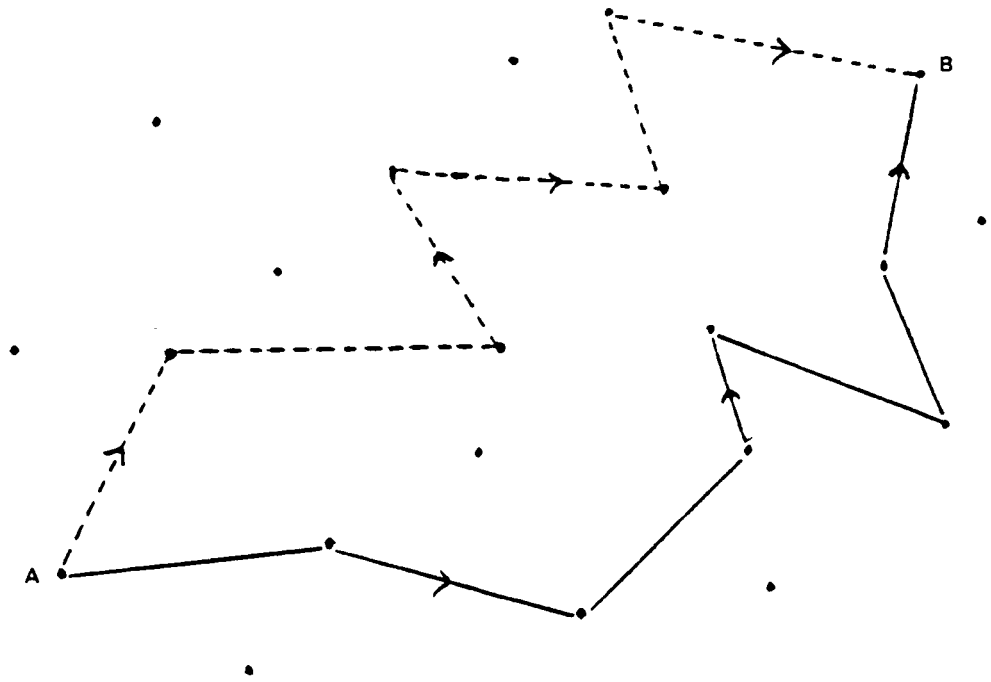


Figure 4.2: Two possible paths for a particle to diffuse from A to B.

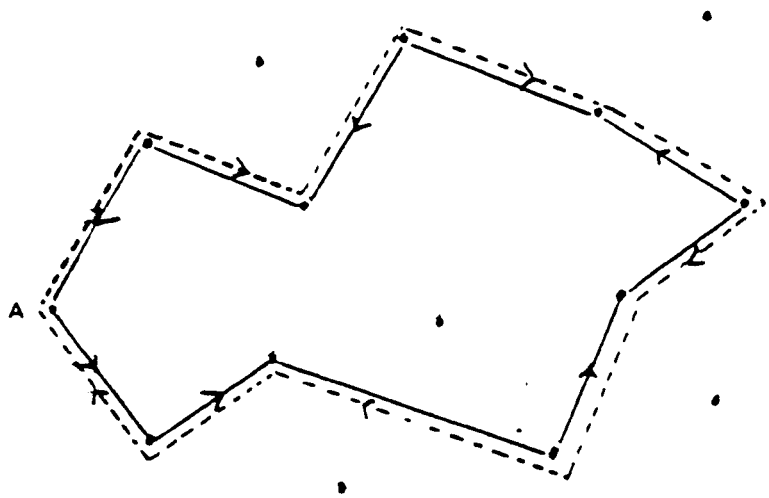


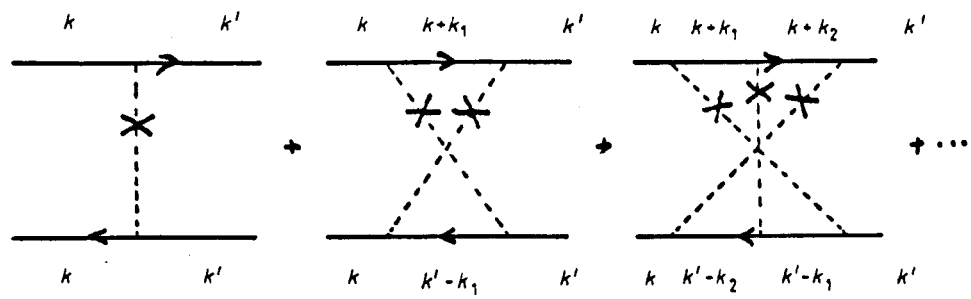
Figure 4.3: Two paths contributing to the backscattering process involving the same loop of impurities. In the presence of time reversal symmetry, the amplitude for each path is the same and so they add constructively.

large number of paths, will usually be zero. This is the assumption that Boltzmann theory makes. There is however one very important case where this need not necessarily be true. Consider the probability for a particle to return to its initial starting point (figure 4.3), and in particular, consider the two paths shown where the amplitudes A_1 and A_2 correspond to a passage around a closed loop of impurities but in opposite senses. In this case, A_1 and A_2 interfere constructively so we must write,

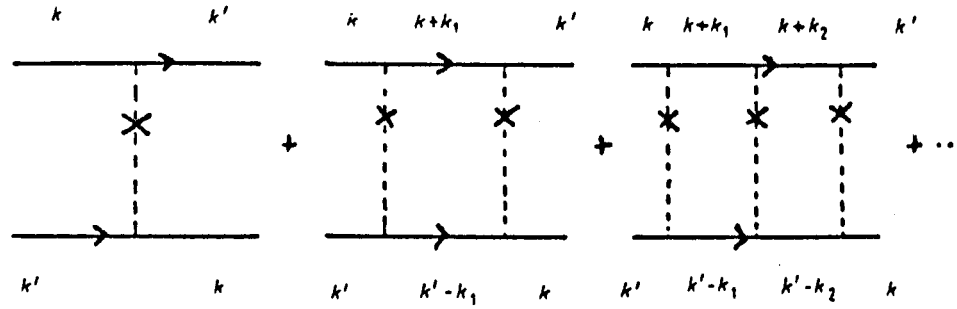
$$|A_1+A_2|^2 = |A_1|^2 + |A_2|^2 + 2\mathrm{Re}A_1A_2^* = 4|A_1|^2 \qquad (4.12)$$

The classical result neglects the interference term and so predicts the probability of return is $2|A_1|^2$. Providing we can show they have a finite weight, inclusion of these interference terms will therefore mean there is a greater possibility of returning to the starting point (backscattering) and so the resistance of a sample will increase. This physical picture emphasizes the quantum nature of the localisation phenomenon, which leads to a negative correction term to the Boltzmann conductivity.

When we expanded the S-matrix to calculate the conductivity (equation 3.14) an infinite number of diagrams were produced. One subset gave rise to the Boltzmann result. We anticipate that another subset will represent the leading correction to σ_B and will have its origins in weak localisation. These diagrams are the maximally crossed diagrams, first discussed in a different context by Langer and Neal (1966), Neal (1967) and in detail by Gor'kov et al (1979) and Abrahams et al (1980),



They account for precisely the kind of interference terms we have just been considering, but represented in momentum rather than real space (Bergmann 1983). A customary redrawing is to reverse the lower line which gives the ladder diagrams in the particle-particle channel (known as the Cooperon),



Each diagram is of the same order so the entire series has to be summed. For zero-range scatterers, the diagrams form a geometric series and so the summation is trivial. The interesting thing is that the sum diverges for small Q, ω* (where $Q = k+k'$) like $(DQ^2 - i\omega)^{-1}$. D is the diffusion constant $= v_F^2 \tau/d$. This strong correlation between states of opposite momenta is the signature of localisation; naively if one wanted to make a localised wavepacket, one would add together plane waves of opposite momenta to make standing waves. This idea that correlation between states of opposite momenta is responsible for localisation incidently disagrees with the model of Kaveh and Mott (1981), and Berggren (1982a),b)) which is now (for this reason) believed to be incorrect (Davies 1987, private communication).

The divergence arises because in an infinite system there are arbitrarily many loops which can contribute to the backscattering process. In a finite system however the maximum loop size is of the order of the system size L, which introduces a lower cut-off in the Q – summation of order $1/L$. The correction to the conductivity is then (for $\omega=0$),

$$\sigma_{WL} = -\frac{2e^2}{\hbar\pi}\frac{1}{L^d}\sum_Q \frac{1}{Q^2} \tag{4.13}$$

where the upper Q cut-off is $1/l$. Notice that the correction term is both negative and size dependent. In $3D$,$2D$ and $1D$ this predicts (assuming that $L \gg l$),

$$\sigma_{3D} = \sigma_B - \frac{e^2}{\hbar\pi^3}\left(\frac{1}{l} - \frac{1}{L}\right) \tag{4.14a}$$

$$\sigma_{2D} = \sigma_B - \frac{e^2}{\hbar\pi^2}\ln(L/l) \tag{4.14b}$$

$$\sigma_{1D} = \sigma_B - \frac{e^2}{\hbar\pi}L \tag{4.14c}$$

The weak localisation correction depends only upon simple constants and so is the same for all materials, whether they are semiconductors or metals. This agrees with the predictions of scaling theory. As the length scale L increases, the $3D$ conductivity saturates to a length independent

*ω refers to the frequency of the external applied field.

constant as we anticipated would happen above the Metal-Insulator transition. In $2D$ and $1D$, the weak localisation term grows until it dominates the Boltzmann term (at which point perturbation theory breaks down). From the length scale at which $\sigma_{WL} \approx \sigma_B$ we can deduce the approximate size of the localisation length. We find that in $2D$ $\xi_{loc} = l\exp(\pi k_F l/2)$ (so if $k_F l$ is large then the localisation length is also), whilst in $1D$, $\xi_{loc} = \pi l$, and all states are localised on a length scale of the mean free path (Thouless 1973).

So far the discussion has been restricted to zero temperature. Before discussing finite temperature, recall the nature of the singularity in the Cooperon at finite frequencies. Had we retained ω, then the length scale $L_\omega = (D/\omega)^{1/2}$ would have provided a natural cut-off in the $Q-$ summation, and would (if $L_\omega < L$) have replaced L in the above formulae (equation 4.14a),b),c)). This is extremely significant for it is saying that finite frequencies can cut-off the weak localisation phenomenon (Gor'kov et al 1979, Al'tshuler et al 1982). If the loop size (figure 4.3) is greater than L_ω, the A.C. field destroys the phase coherence needed for localisation and so provides the length scale which determines the magnitude of the weak localisation term. To introduce finite temperatures, we make an analogy with the above and note that inelastic scattering will also destroy the phase coherence necessary for localisation. If the inelastic effects are characterised by a time τ_ϕ, the length scale $L_\phi = (D\tau_\phi)^{1/2}$ will then be the length scale which replaces L. Thouless (1977) interpreted the time τ_ϕ as the lifetime of an eigenstate of the disordered system. Al'tshuler et al (1982) however pointed out the fact that the significant timescale is the time over which phase coherence is lost, and called τ_ϕ the phase relaxation time. The phase relaxation time need not be the same as the the inelastic time.

The phase relaxation time τ_ϕ is temperature dependent, and at low temperatures is controlled by electron-electron scattering. When low momentum transfers are important, the phase relaxation time is the same as the electron-electron scattering time, which in disordered systems varies as $T^{-d/2}$ (Schmid 1974). When low energy transfers are important, the relevant process is equivalent to the interaction of an electron with fluctuating electromagnetic fields, the so called Nyquist mechanism. The temperature dependence of τ_ϕ in this case goes as $T^{2/(d-4)}$ (Al'tshuler et al 1981, 1982). The sensitivity of the weak localisation term means that it controls the

temperature dependence of the conductivity, since at low temperatures, σ_B is essentially independent of T. As the temperature drops, so inelastic effects become less important, causing the weak localisation correction to σ to increase and the resistance to increase. In $2D$ the prediction is that there will be a $\ln(T)$ correction to the resistance. This was first observed by Dolan and Osheroff (1979) and has been reproduced many times (see the reviews by Bergmann 1984, Lee and Ramakrishnan 1985). In one dimensional wires it is more difficult to interpret the data, though there are now reports of the observation of the Nyquist mechanism, with a $T^{-1/3}$ dependence found in the length L_ϕ (Thornton et al 1986, Lin and Giordano 1987).

4.5 Magnetic fields and interaction effects.

The inherent instability of the Cooper channel to perturbations which destroy phase coherence (violate time reversal symmetry) means many other effects can introduce new length scales into the problem. In the presence of a magnetic field, the change of phase gained by going around a loop of impurities depends upon the sense in which the loop is traversed. As a result, the relative phase upon recombination depends upon the loop size and so on average, magnetic fields will destroy localisation, and lead to negative magnetoresistance. The theory was first worked out for $3D$ systems by Kawabata (1980,1981), and for $2D$ systems in a perpendicular field by Al'tshuler et al (1980). The extension to longitudinal fields in a $2D$ system of width a, as well as cylindrical wires with either field orientation was made by Al'tshuler and Aronov (1981). The behaviour found in these cases depends subtly on the relative magnitudes of the magnetic length L_H , L_ϕ and the channel width a. Experimental results are in good agreement (for example Newson and Pepper 1985, Newson et al 1985, Thornton et al 1986, Lin and Giordano 1987). Magnetic fields can also lead to very interesting effects in multiply connected structures, with both oscillations in ϕ_0 (the flux quantum) and $\phi_0/2$ being predicted and observed (Al'tshuler et al 1981, Sharvin and Sharvin 1981, Datta and Bandyopadhyay 1987, Dolan et al 1986, Washburn et al 1987). Many of the relevant experiments are described in the recent review by Aronov and Sharvin (1987).

Spin-orbit scattering defines a new timescale τ_{so} and a new length scale L_{so}. Hikami et al (1980) considered the behaviour of a system at finite temperature and magnetic field, when spin-

orbit scattering is present. The magnetoresistance can now have quite exotic behaviour and may be both positive and negative. The spin-orbit effect is greatest in metals, and the theory has been confirmed to quite remarkable accuracy (Bergmann 1982, Santhanam et al 1984, Licini et al 1985, Lin and Giordano 1987). Another mechanism which can produce similar effects is spin-flip scattering due to magnetic impurities (Bergmann 1984).

So far we have said little about electron-electron interactions. In addition to localisation, which is a single particle phenomenon, interaction effects have also been studied extensively in disordered solids. In the strongly localised regime, the theory is usually developed in terms of various parameters (such as the Hubbard U) which characterise the electron-electron repulsion term. It is believed that interaction effects can produce a similar effect to the Metal-Insulator transition (the Mott transition) by opening up a gap in the density of states at the Fermi level (the coulomb gap). For detailed reviews of these and other phenomena in the strongly disordered regime see the reviews by Pollak and Ortuno 1985 and Efros and Shklovskii 1985. In the weakly localised regime, more is known qualitatively because diagrammatic perturbation theory is applicable. The important point is that because the electron motion is diffusive rather than ballistic, the electron-electron interaction is generally enhanced. It is known for example that interactions lead to a $\ln T$ term in the conductivity in $2D$, exactly as the localisation term does. It is also known that interaction effects are much less sensitive to weak magnetic fields than localisation effects are, and so it is possible to distinguish the two contributions by suppressing the weak localisation term in a magnetic field. To a first approximation, the two effects are additive. It is only because of this that it makes sense to talk of the localisation problem in isolation. Excellent reviews of interaction effects in the weakly disordered regime have been given by Al'tshuler and Aronov 1985, Fukuyama 1985 and Lee and Ramakrishnan 1985.

4.6 The one and quasi-one-dimensional regime.

As we have been interested in systems which are inherently one dimensional we shall study this regime in a little more detail. This will highlight some of the important ideas we wish to return to later in this work. Since Mott and Twose (1961), further investigations have proceeded

along two distinct paths. The first route has relied on the Green's function methods; essentially solving a kinetic equation. Berezinskii (1973) considered the disorder to all orders to prove that all the states in one dimension are indeed localised. This view was also supported by Abrikosov and Rhyzkin (1978). For wires in the weak localisation regime, Al'tshuler and Aronov (1985) argued that the corrections to the conductance will be strictly one dimensional if the width of the wire is less than the relevant length scale, otherwise they will be two or three dimensional depending on the geometry. In the absence of magnetic fields they suggested that the relevant length scale is the phase breaking length L_ϕ. They also considered finite length effects (Al'tshuler and Aronov 1984) and showed that for very short wires ($L \sim L_\phi$) boundary effects in the longitudinal direction are also important. This idea has been developed by Santhanam (1987), and the effect has been observed experimentally by Masden and Giordano (1984).

The second approach to the problem has utilised the Landauer formula (1970), or variations of it. Landauer expressed the conductance of a one-dimensional system in terms of the transmission probability T for an incident particle to penetrate a length L,

$$G(L) = \frac{e^2}{\pi\hbar} \frac{T(L)}{1-T(L)} \tag{4.15}$$

and showed that the ensemble average $<G> \propto e^{-L/\xi}$. Clearly this implies localisation of the states as $L \to \infty$. The derivation of equation (4.16) is sufficiently simple to make it an attractive method for studying transport in disordered solids. (A word of warning; although the derivation is simple it is also subtle, and attempts to generalise the result to higher dimensions lead to controversy for a number of years. This has now been satisfactorily resolved; for a chronological development of the problem see Landauer 1970, Economou and Soukoulis 1981, Fisher and Lee 1981, Langreth and Abrahams 1981, Thouless 1981, Economou and Soukoulis 1981, Buttiker et al 1985, Eranen and Sinkkonen 1987). A most important point is that the Landauer formula can be applied to a single sample rather than representing some kind of ensemble average. This has revealed a whole range of new phenomena in small systems which are not seen after averaging has taken place. In the strongly localised regime, the conductance is a violently fluctuating function of Fermi energy in 1D (Azbel and Soven 1983, Lee 1984, McInnes and Butcher 1985). For a given Fermi energy,

the conductance will also vary enormously from sample to sample, and as a result $<G>$ is not a statistically well defined quantity. This reflects the fact that the mean conductance is very different to the typical conductance in $1D$. Anderson et al (1980) used the Landauer formula to show that $\ln(1+g^{-1})$ is the well defined scaling variable in that it has additive mean. This idea was pursued in a second paper by Anderson (1981) and a heuristic generalisation to higher dimensions given. Using these higher dimensional type expressions, the conductance of finite cross-sectional wires has been discussed by Azbel (1980a,b, 1981), small rings by Buttiker et al (1985) and even device characteristics have been calculated (see e.g. Davies 1987). The Landauer formula is ideal for performing numerical calculations of this type. On the other hand it is not so easy to extract analytical results, and the extension to finite temperatures is not trivial. Nevertheless it seems that the Landauer expressions are becoming more widely used in studying small systems, especially when the scattering is small or even non-existent (Van Wees et al 1988).

4.7 Fluctuation phenomena.

The Landauer picture tells us that the conductance of small samples is sample dependent. Experiments in ring geometries and in wires have revealed such sample dependent features as a function of magnetic field (Umbach et al 1984) and chemical potential (Kaplan and Hartstein 1986). The important feature of these fluctuations is that they are reproducible in a given device and of universal amplitude at $T=0$. The samples they are observed in are metallic ($L \gg \xi$) and are called mesoscopic, meaning somewhere between macroscopic and microscopic. The fluctuations in mesoscopic devices (Universal Conductance fluctuations) are now relatively well understood (Al'tshuler 1985, Lee and Stone 1986, Al'tshuler and Khmel'nitskii 1986, Lee et al 1987).

The Kubo formula, in principle, gives us the conductance of a specific sample with a specific impurity configuration. In practice this is too difficult to evaluate which is why an ensemble average is taken. In order to study sample specific fluctuations therefore, we need a way of getting around this problem. Lee and Stone (1985) hypothesised that the fluctuations in the conductance of a given sample as a function of Fermi energy or magnetic field could be viewed as the statistical fluctuations which would occur in an ensemble of samples differing only in their

microscopic impurity configurations. This enabled ensemble averaging and standard techniques to be reintroduced into the problem. In particular, they considered the correlation function,

$$F(\varepsilon_F, \Delta\varepsilon, B, \Delta B) = \langle g(\varepsilon_F, B)\, g(\varepsilon_F+\Delta\varepsilon, B+\Delta B)\rangle - \langle g(\varepsilon_F, B)\rangle\langle g(\varepsilon_F+\Delta\varepsilon, B+\Delta B)\rangle \quad (4.16)$$

where <....> denotes ensemble averaging over many samples. When $\Delta\varepsilon = \Delta B = 0$, F is simply the variance of the conductance amongst the ensemble of samples for a given ε_F, B. The assumption of the ergodic hypothesis is that this is representative of the magnitude of the fluctuations in a given sample. In the other limit where $\Delta\varepsilon$ and ΔB are large, the conductances will be uncorrelated and F will be zero. The decay lengths E_c, B_c of F will then be representative of the mean spacing between the fluctuation peaks. The correlation function has been evaluated diagrammatically (Lee et al 1987) and the main results are as follows. The amplitude of the fluctuations in ε_F or B is independent of the size of the sample and only weakly dependent upon its shape. Furthermore, this amplitude is universal and of order e^2/h at T=0. The spacing between the peaks in energy is given by,

$$E_c = \hbar\pi^2 D/L_z^2 \quad (4.17)$$

which just relates to the time for the electrons to diffuse the length of the sample in the current carrying direction (Thouless 1977). The corresponding expression for B_c is a little more complicated since it depends upon the orientation of the field (Lee et al 1987). We shall find later that similar ideas can be found for the thermal transport coefficients, which also show fluctuations in mesoscopic devices. The important point to stress is that the origin of these fluctuations is quantum mechanical and not classical. To see this, imagine forming a large conductance by the adding together of smaller conductances which are essentially randomly distributed. Classically the variance of the total conductance is given by the central limit theorem,

$$\frac{var\ (G)}{<G>^2} \sim \frac{1}{L^d} \quad (4.18)$$

and since $<G> \sim L^{d-2}$, this predicts that $var\ (G) \sim L^{d-4}$. In all dimensions less than $d = 4$ the conductance therefore self-averages classically. In the present problem however the variance is independent of system size; the quantum mechanical origin of the fluctuations exactly cancels the classical self-averaging behaviour and the fluctuations persist out to all length scales. This is true for $T = 0$. At finite temperatures however there are two important length scales which enter the

problem and lead to averaging effects. Firstly the work on weak localisation has shown that a system is quantum mechanical up to a length scale L_ϕ whereupon classical combination rules are appropriate. At length scales greater than L_ϕ we can therefore sub-divide our sample into regions of size L_ϕ which are then added classically to give,

$$var\ G(L,T) = var\ G(L = L_\phi, T) \left[\frac{L_\phi}{L}\right]^d \qquad (4.19)$$

As the sample size exceeds the phase breaking length, the fluctuations therefore begin to self average, and die away completely in the thermodynamic limit. The other important length scale is the thermal diffusion length $L_T = (hD/kT)^{1/2}$ which determines energy averaging (thermal broadening) in the sub-regions defined on a length scale L_ϕ. In particular if $L_T \ll L_\phi$ the fluctuations will have been essentially averaged out even at a length scale $L_z = L_\phi$, because in this case $k_B T \gg E_c$. What we are then left with is the average behaviour such as weak localisation. Experiments to look for fluctuations must therefore be performed at suitably low temperatures where the system size is much less than either of these lengths. A very detailed account of all finite temperature effects has been given by Lee et al (1987), and recent reviews have been given by Lee 1987, Al'tshuler 1987 and Webb et al 1987.

CHAPTER 5

WEAK LOCALISATION IN THE QUASI-ONE-DIMENSIONAL REGIME.

5.1 Introduction to the problem.

In the previous chapter, a general overview of localisation and in particular weak localisation was given. We now wish to perform a calculation of the weak localisation term in a quasi-one-dimensional system explicitly retaining the sub-band picture to account for finite width effects. This will accomplish two things. Firstly the results will be expressed in terms of an anisotropic diffusion constant, with differing longitudinal and transverse components. Secondly, since the results are valid for an arbitrary number of occupied sub-bands, and hence arbitrary channel width, it enables an explicit calculation of the dimensional crossover from a $1D$ to a $2D$ scaling. This is particularly interesting since many experiments can now probe such a transition. The role the sub-bands play in determining the weak localisation correction turns out to be rather different from their role in determining the Boltzmann term.

Most of the work that has been carried out on weak localisation has assumed an isotropic diffusion constant. Anisotropy can arise in many ways however so it important to understand how this affects the results. The simplest way it can happen is if the effective mass is anisotropic, as in silicon for example. This case was considered by Wolfle and Bhatt (1984), who found that the effect of the anisotropy can be completely absorbed into an anisotropic diffusion constant,

$$\delta\sigma_{\mu\mu} = -\frac{2e^2}{\pi\hbar} D^0_{\mu\mu} \sum_{\mathbf{q}} \frac{1}{-i\omega + \sum_{\mu} D^0_{\mu\mu} q_\mu^2} \tag{5.1}$$

We shall find a similar result holds when the anisotropy is due to finite width effects (sub-bands), and shall derive analytic expressions for the $D^0_{\mu\mu}$ for this case. The correction term always has the same anisotropy as the Boltzmann term. The consequences of equation (5.1) have been investigated experimentally and found to be correct (Bishop et al 1984).

When sub-bands were discussed earlier, a general feature was the lack of translational symmetry in the confined directions leading to Green's functions non-diagonal in the sub-band indices. The problem was simplified by assuming that the sub-bands were sufficiently far apart

for the off-diagonal elements to be small. If on the other hand we want to demonstrate a $1D$ to a $2D$ crossover where near the $2D$ limit the sub-bands are necessarily almost degenerate, this approximation is no longer good enough. In order to proceed we shall therefore use a slightly different model which assumes periodic boundary conditions in all directions (Fisher and Lee 1981, Cantrell and Butcher 1985). The repeat symmetry in the confined directions produces a Green's function which is diagonal in the sub-band labels without recourse to approximation once the ensemble average is taken. The choice of such a model is not expected to affect the essential physics in the problem (a point we shall return to) and is certainly sufficient for illustrative purposes. For a non-interacting electron gas confined to a region ($L \times W$) the wavefunctions in this model are,

$$\phi_{\alpha k} = \frac{1}{(LW)^{\frac{1}{2}}} e^{ikx} e^{i\alpha z} \tag{5.2}$$

where k is taken to be continuous and,

$$k_\alpha = 2\pi\alpha / W \qquad (\alpha = 0, \pm 1, \pm 2, \cdots)$$

α thus plays the role of a sub-band index, and the unperturbed sub-band origins are simply $(2\pi/W)^2 \alpha^2/(2m)$. (In the interests of conciseness, we set $\hbar=1$ and $m^*=m$, and restore the full dependences at the end of the calculation). The thickness in the third dimension will be taken to be sufficiently small that the electrons are always in the ground sub-band associated with confinement in that direction. Consequently it can be ignored altogether without loss of generality. This situation is relevant to pinched MOSFETS for example. When the width W is small, the k_α are discrete and represent the different sub-bands. In the opposite limit, the k_α are quasi-continuous and the wavefunction is simply a $2D$ plane wave. This simple crossover in the wavefunction will reflect the crossover in the transport coefficients we wish to explore.

5.2 The single-particle self-energy.

The same model for the impurities will be used as in previous chapters. Within the present model the matrix elements (squared) are simple constants, and so the self-energy is only a function of the energy variable ε. In order to ensure the correct divergence in the Cooperon (which is responsible for the weak localisation behaviour and guaranteed in the presence of time reversal

symmetry), it is important that the self-energy is treated consistently with the approximations made in evaluating the current-current correlation function. The divergence arises as a result of a delicate cancellation between these one and two particle properties. We will show that the self-consistent Born approximation considered earlier is appropriate. Again we will ignore the real part of the self-energy. (Retaining the real part throughout leads to the same divergence in the Cooperon and only alters the single-particle properties). If $-\Gamma(\varepsilon)$ is the imaginary part of the retarded self-energy then,

$$-\Gamma(\varepsilon) = \mathrm{Im}\left[\frac{n_i U_0^2}{LW}\right] \sum_{q k_\beta} (\varepsilon+\mu-q^2/2m-k_\beta^2/2m+i\Gamma(\varepsilon))^{-1} \tag{5.3}$$

which, since Γ is independent of momentum or sub-band labels is equivalent to saying,

$$1-\left[\frac{n_i U_0^2}{LW}\right] \sum_{q k_\beta} G^R(q, k_\beta, \varepsilon)\, G^A(q, k_\beta, \varepsilon) = 0 \tag{5.4}$$

where $G^{R,A}$ are the retarded and advanced Green's functions respectively.* As well as determining Γ this equation provides a consistency requirement which must be maintained throughout the calculation. Since q is a quasi-continuous, one-dimensional variable, the sum over q may be performed by a contour integral, with the result that,

$$\Gamma(\varepsilon) = \frac{n_i U_0^2 m}{W} \sum_{k_\beta} \frac{a}{a^2+b^2} \tag{5.5}$$

where the identity,

$$(a+ib) = (2m(\varepsilon+\mu)-k_\beta^2+i2m\Gamma)^{1/2}, \qquad b>0 \tag{5.6}$$

defines a and b. As it stands, equation (5.6) is still a complicated expression for Γ since a and b depend upon Γ themselves, and in general must be solved numerically. The density of states in this model is linearly related to Γ,

$$g(\varepsilon) = \frac{2LW\Gamma}{n_i U_0^2 \pi} \tag{5.7}$$

and when the sub-bands are well separated, will show structure similar to that discussed previously (figure 3.2).

The various limiting behaviours of Γ serve as a demonstration of dimensional crossover in multi-sub-band transport. The separation of the sub-bands is controlled by the system width W. If W is sufficiently narrow so that the broadening is less than the sub-band separation energy,

* $G^{R,A}(q, k_\beta, \varepsilon) = (\varepsilon+\mu-q^2/2m-k_\beta^2/2m \pm i\Gamma(\varepsilon))^{-1}$

then the sub-band concept is meaningful and they are well defined. As the width of the channel increases, so the sub-band separation decreases (as $1/W^2$) and eventually Γ will be of the order of this spacing. At this point, the concept of sub-bands becomes meaningless and the self-energy rapidly approaches its two-dimensional limit. As an order of magnitude estimate of this effect, suppose the system is in the $2D$ limit where $\Gamma \propto n_i U_o^2 m$. As we decrease the width W, Γ will stay roughly at this value until the sub-band spacing is of same order. If the Fermi energy at this point lies between the nth and $(n+1)th$ sub-bands this is equivalent to saying,

$$\Gamma_{2D} \simeq n_i U_o^2 m \simeq [(n+1)^2 - n^2] \frac{\hbar^2 4\pi^2}{2m^* W^2}$$

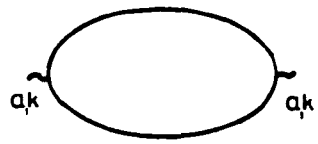
Now since $k_F \simeq 2\pi n/W$, this implies,

$$W^2 - 4\pi W l - 4\pi^2 l/k_F = 0$$

where l is the $2D$ mean free path $= \hbar^2 k_F / 2m\Gamma$. In the weakly disordered regime where the theory is applicable, $k_F l \gg 1$ and so $W \simeq 4\pi l$. This is an important result. It states that if we define a narrow channel in a $2DEG$, then the sub-band structure only becomes apparent for widths less than the mean free path in the $2DEG$. At length scales smaller than this the density of states will begin to show structure and the possibility of observing QSE becomes realistic. At greater length scales the density of states is simply a constant, exactly what we would expect of the $2D$ limit. Although this calculation has been performed for quadratically spaced sub-bands, we expect that the essential result (that the relevant length scale is the mean free path) is the correct one for the density of states, even if the sub-bands are linearly separated in energy for example.

5.3 The weakly localised conductance.

Within this new model, we proceed in the usual fashion and expand the current-current correlation function using the $S-$ matrix. In the previous chapter we saw that the relevant quantity to discuss is the conductance rather than the conductivity, and so even though the difference is trivial in this limit we shall refer to conductances from now on. The Boltzmann term we associate with the simple bubble diagram,

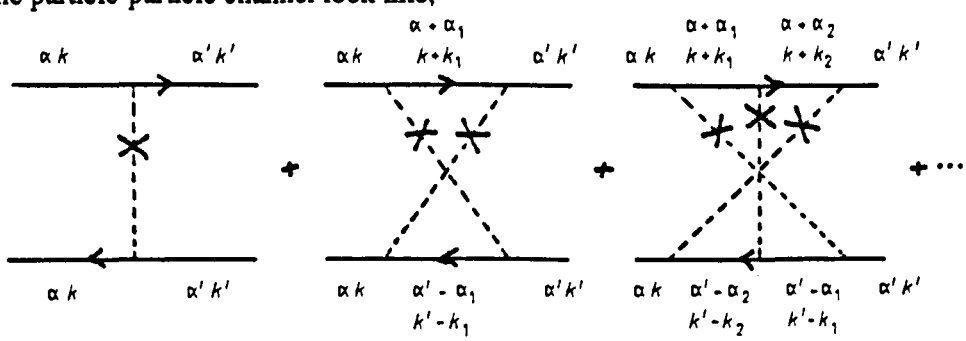


In chapter 3 this diagram was evaluated in detail (equation (3.21)). Performing the sum over k by a contour integral gives (at T=0),

$$G_B = \frac{e^2}{\pi m \Gamma L} \sum_{i} \frac{a^3}{(a^2+b^2)} \tag{5.8}$$

where a and b are defined through equation (5.6) and evaluated at ε=0. This expression is the zero temperature Boltzmann conductance for an arbitrary number of occupied sub-bands. When the sub-bands are well separated in energy this will show oscillations as a function of Fermi level, which is simply the manifestation of the quantum size effects discussed in detail in chapter 3. When the channel width is wide, the sub-bands mix together and G_B tends to its 2D limit. The criterion for this to happen is exactly the same as for the self-energy and the density of states, namely that the channel width should be of the order of the mean free path. Above that the Boltzmann conductance will appear two-dimensional and no QSE are to be expected. Experiments to resolve QSE must therefore take place in sufficiently narrow channels of course.

The weak localisation term is associated with the maximally crossed diagrams, which when drawn in the particle-particle channel look like,



We use the notation Q=k+k' and analogously K=k_α+k'_α. As usual, the choice of δ– function scattering potentials enables the series to be summed as a geometrically. The important contribution to the weak localisation comes from the terms where the Green's functions have the opposite

analyticity i.e. when the upper line is retarded and the lower advanced. Retaining these terms then gives,

$$\Delta G_{WL} = -\frac{e^2}{\pi m^2 L^2} \int_{-\infty}^{\infty} d\varepsilon \sum_{k,k_\alpha,Q,K} k(-k+Q)\, G^R(k,k_\alpha,\varepsilon)\, G^A(k,k_\alpha,\varepsilon)\, S(Q,K,\varepsilon)$$
$$\times G^R(-k+Q,-k_\alpha+K,\varepsilon)\, G^A(-k+Q,-k_\alpha+K,\varepsilon)\, \frac{\partial f(\varepsilon)}{\partial \varepsilon} \qquad (5.9)$$

The scattering vertex may be written in the simple closed form,

$$S(Q,K,\varepsilon) = \frac{\Gamma_0}{1-\Gamma_0 R} \qquad (5.10)$$

where $\Gamma_0 = n_i u_o^2/WL$ and,

$$R = \sum_{q,k_\beta} G^R(q,k_\beta,\varepsilon+i\omega)\, G^A(-q+Q,-k_\beta+K,\varepsilon) \qquad (5.11)$$

We have retained the finite frequency component ω in the scattering vertex as a guide to how the phase relaxation time will enter the problem. Equation (5.9) is the full expression for the weakly localised correction to the Boltzmann conductance. Notice that it contains the factor $\partial f/\partial\varepsilon$ which is responsible for thermal broadening. Thermal broadening is usually considered unimportant in the theory of weak localisation, but in sub-band systems may lead to sizeable effects, and its importance in certain regimes will be discussed in a little while.

In the limit $Q,K,\omega\to 0$ the scattering vertex diverges (compare the denominator of equation (5.10) with the R.H.S. of equation (5.4) which is identically zero). This divergence arises because we have been careful in maintaining self-consistency between the two particle correlation function and the single particle self-energy. Since the small Q,K,ω terms are therefore going to be the most important, it is admissible to work only to lowest order in these quantities, which means we can neglect all Q,K dependence except that arising in the scattering vertex. To begin with we shall work at $T=0$. The sum over k in equation (5.9) may be performed by a contour integral (with two double poles in the upper half plane) with the result,

$$\Delta G_{WL} = -\frac{e^2}{2\pi m L \Gamma^3} \sum_{k_\beta} \frac{a^3}{(a^2+b^2)} \sum_{Q,K} S(Q,K,\varepsilon=0,\omega) \qquad (5.12)$$

where a and b are defined as before. Next we perform the sum over q in equation (5.11) for R, again by a contour integral (but with only two single poles in the upper half plane). The answer is rather complicated, however since we are only interested in small values of its arguments we can

expand to lowest order to give,

$$R \simeq \sum_{k_\beta} \frac{Lma}{\Gamma(a^2+b^2)} \left[1 - \frac{Q^2a^2}{4m^2\Gamma^2} - \frac{K^2k_\beta^2a^4}{4m^2\Gamma^2(a^2+b^2)^2} + \frac{i\omega}{2\Gamma} \right] \tag{5.13}$$

In order to facilitate the derivation one makes use of the symmetry properties of the original expression. The leading term of $\Gamma_0 R$ is unity (c.f. equation 5.6) ensuring the correct divergence in the Cooper channel. To this order, S may be written,

$$S(Q,K) = \frac{2\Gamma n_i U_o^2}{WL} \frac{1}{DQ^2 + D_p K^2 - i\omega} \tag{5.14}$$

where D and D_p are defined by,

$$D = \frac{n_i U_0^2}{2Wm\Gamma^2} \sum_{k_\beta} \frac{a^3}{(a^2+b^2)} \tag{5.15a}$$

$$D_p = \frac{n_i U_0^2}{2Wm\Gamma^2} \sum_{k_\beta} \frac{k_\beta^2 a^5}{(a^2+b^2)^3} \tag{5.15b}$$

Notice that exactly the same combination of factors occurs in equation (5.15a) for D as in the first factor of equation (5.12) for ΔG_{WL}. The net result is that we can write,

$$\Delta G_{WL} = -\frac{2e^2D}{\pi L^2} \sum_{Q,K} \frac{1}{DQ^2 + D_p K^2 - i\omega} \tag{5.16}$$

Compare the above result with the result of Wolfle and Bhatt (1984) (equation 5.1). The two results are identical and so we identify D and D_p with the longitudinal and transverse diffusion coefficients in a multi-sub-band system. The anisotropy in the present case is not due to an anisotropic effective mass but due to finite width effects. Let us consider the form of these diffusion constants a little more closely. Suppose the sub-bands are well separated and that all the electrons are confined to lowest sub-band. It is easy to show in that in this case $\Gamma \simeq n_i U_0^2/Wk_F$, $a \simeq k_F$ and $b \simeq 0$ whereupon $D = k_F^2 \tau/m^2$ and $D_p \simeq 0$. Thus in this limit we have the usual $1D$ diffusion down the channel and almost no diffusion across it. In the opposite limit when W is in excess of the $2D$ mean free path so that many sub-bands are occupied, we can replace the sum over k_β by an integral to evaluate the diffusion coefficients. This is rather trivial and gives us the result that we would expect, namely that both D and D_p tend to the $2D$ result ($k_F^2 \tau/2m^2$) and we regain the isotropic situation. Thus our interpretation of these constants as describing the diffusion in a multi-sub-band channel is very plausible. As further evidence, it should be noted that D

is related exactly to the Boltzmann conductance (equation 5.8) and the density of states (equation 5.7), via the usual Einstein relation.

5.4 The one dimensional limit for ΔG_{WL}.

The expression we have derived for the weak localisation correction to the conductance is valid as long as the expansion in Q, K is valid. This implies that the terms in equation (5.13) involving Q and K should both be less than unity for all values of k_β. There are three regimes to consider, when $|k_\beta| \ll k_F$, $|k_\beta| \gg k_F$ and $|k_\beta| \simeq k_F$, and we want to ensure that $Q < 2m\Gamma/a$ and $K < 2m\Gamma(a^2+b^2)/(k_\beta^2\, a^4)$. The results are presented in tabular form below:

REGIME	VALUE OF a	CONSTRAINT ON Q	CONSTRAINT ON K
$\lvert\kappa_\beta\rvert \ll \kappa_f$	$a \simeq k_f$	$Q < 2m\Gamma/\kappa_f$	$K < 2m\Gamma/\kappa_\beta$
$\lvert\kappa_\beta\rvert \gg \kappa_f$	$a \simeq m\Gamma/\kappa_\beta$	$Q < 2\kappa_\beta$	$K < \kappa_\beta^3/m\Gamma$
$\lvert\kappa_\beta\rvert \simeq \kappa_f$	$a \simeq (m\Gamma)^{1/2}$	$Q < 2(m\Gamma)^{1/2}$	$K < 4m\Gamma/\kappa_f$

All these conditions are satisfied if Q and K are both restricted to be less than $2m\Gamma/k_F$, which then constitutes an upper cut-off to the Q, K summation. This quantity is simply $1/l$ where l is the elastic scattering length at the Fermi surface. This is the same criterion as used in conventional treatments for the upper cut-off, which is not that surprising as l represents the smallest loop size (figure 4.3) that can yield the interference necessary for weak localisation.

To begin with we shall consider only the K=0 term. In the standard fashion inelastic effects are introduced by replacing $i\omega$ by $1/\tau_\phi$ where τ_ϕ is the phase relaxation time at the Fermi level. Replacing the sum over Q by an integral we can easily derive (reinserting $\hbar$),

$$\Delta G_{WL} = -\frac{2e^2}{\pi^2\hbar L} L_\phi \tan^{-1}\frac{L_\phi}{l} \tag{5.17}$$

where $L_\phi = (D\tau_\phi)^{1/2}$ and $l \simeq (D\tau)^{1/2}$. Since at low temperatures $L_\phi \gg l$ we finally obtain,

$$\Delta G_{WL} = -\frac{e^2}{\pi\hbar L} L_\phi \tag{5.18}$$

This is the exact one dimensional result. Thus in situations where the $K = 0$ term dominates, regardless of the number of occupied sub-bands, the weak localisation scaling will be linear in L_ϕ.

and all the information about the sub-bands is contained in L_ϕ. A simple qualitative insight into this is provided by recognising that weak localisation is a quantum interference phenomenon. $K = 0$ implies an interference between states in "opposite" sub-bands (i.e. k_α and $-k_\alpha$) which (in this model) in some sense corresponds to transverse standing waves. The resulting contribution thus arises solely from motion along the strip and appears one-dimensional. Notice that the length L_ϕ depends upon the longitudinal diffusion constant. This is the length scale which enters the formula for the conductivity. The length scale which controls the dimensionality of the correction term on the other hand is slightly different, which we shall now explore.

5.5 The intermediate regime and the 2*D* crossover.

We have just shown that the $K = 0$ term is responsible for the weak localisation term behaving one-dimensionally. The question we need to answer therefore is when is this term dominant? Firstly we note that the expansion employed to derive equation (5.16) was only valid for $K<1/l$ and that $K=2\pi n/W$. If $W<l$ this can only be satisfied by $K=0$. This then is the first important result; if the width of the channel is less than the elastic scattering length then the weak localisation correction should appear one-dimensional. From equation (5.16), if $W > l$ the $K = 0$ term will still continue to dominate provided,

$$\left[\frac{2\pi}{W}\right]^2 \frac{D_p}{D} > \frac{1}{L_\phi^2}$$

or,

$$W<2\pi L_\phi\left[\frac{D_p}{D}\right]^{1/2} \tag{5.19}$$

The first remark about this is if the diffusion constant is isotropic (i.e. $D_p = D$) the condition for 1*D* behaviour is that $W<L_\phi$ (the factor of 2π is not that important to our discussion). This was the case considered by Al'tshuler and Aronov (1981) and the result they obtained. We have seen that D_p and D are essentially equal to each other when the channel is very wide, and so this criterion is valid in this limit (the so called dirty channel case). If on the other hand W and hence the number of occupied sub-bands is small, then the full condition above should be used. We can analyse this in a slightly different way by remembering that $L_\phi=(D\,\tau_\phi)^{1/2}$ so,

$$W < 2\pi(D_p \tau_\phi)^{1/2} = L_{tr} \tag{5.20}$$

The length scale L_{tr} is a transverse inelastic length involving the transverse diffusion constant but the same phase relaxation time as in L_ϕ. These two length scales are than related by $L_{tr}/L_\phi = (D_p/D)^{1/2}$. This transverse length is the length scale which controls the system dimensionality as far as weak localisation is concerned, and is our second important result. Combining the two, we predict that the weak localisation in a multi-sub-band wire will scale one-dimensionally if,

$$W < MAX(L_{tr}, l) \tag{5.21}$$

We will discuss the implication for experiment in a moment. The physical significance of the length scale L_{tr} is rather clear, it is simply the distance an electron can diffuse across the channel before suffering a phase destroying scattering event. If W is less than this length, then the electrons can "sense" the confinement which in turn leads to a one-dimensional correction term. We should stress that the distinction between the transverse and phase breaking lengths should be made for any source of anisotropy, and so the same line of argument holds for the effective mass anisotropy discussed at the beginning of this chapter.

When W greatly exceeds either of these lengths we should expect to recover the $2D$ limit. To see this we note that if W is large then $D_p \simeq D$ and $L_{tr} \simeq L_\phi$, and the sum over Q, K may be replaced by integrals. After a simple change of variables we obtain from equation (5.16),

$$\Delta G_{WL} = -\frac{e^2}{2\pi^3\hbar} \int_0^{2\pi} d\theta \int_0^{1/l} \frac{Y\, dY}{Y^2 + 1/L_\phi^2} = -\frac{e^2}{\pi^2\hbar} \ln \frac{L_\phi}{l} \tag{5.22}$$

which is the usual $2D$ result. The full, general expression for the correction is obtained by integrating over Q and leaving the summation over K,

$$\Delta G_{WL} = -\frac{2e^2}{\pi^2\hbar L} \sum_K \frac{L_\phi}{(L_{tr}^2 K^2 + 1)^{1/2}} \tan^{-1} \frac{L_\phi}{l} \left[\frac{1 - l^2K^2}{1 + L_{tr}^2 K^2} \right]^{1/2} \tag{5.23}$$

and is valid for an arbitrary channel width W. The K=0 term is obviously just the one-dimensional result, and the other K terms (if present) represent corrections to it. It is easy to show that at length scales in excess of L_{tr} and l the weak localisation term crosses fairly rapidly and smoothly to its $2D$ scaling in L_ϕ. This is important because it suggests that in the majority of experiments the data can be fitted to either a $1D$ scaling or a $2D$ scaling without having to bother

about the exact functional form in the intermediate regime, according to the criteria for dimensional crossover already given. In principle, equation (5.23) could, however, be used for a more exact numerical simulation of the general weak localisation scaling with L_ϕ.

5.6 Comparison to and implications for experiment.

We have not so far considered the effects of thermal broadening (if any). Thermal broadening can be included by reintroducing the integral over the derivative of the Fermi function in any of the above expressions for ΔG_{WL}. Whether it is important or not depends upon the magnitude of $k_B T$ compared to the sub-band separation energy ε_0. (The experiments are all carried out at temperatures far below the Fermi temperature so we don't have to worry about the size of $k_B T$ in relation to ε_F). We have seen that if the sub-bands are well separated, then quantities may vary with energy on the scale ε_0 and that thermal broadening can be important in such situations. The effect of this can be estimated as follows. If $k_B T \ll \varepsilon_0$ then the broadening is unimportant, and the above expressions are correct with all quantities evaluated at the Fermi surface. The other limit however is a little more complicated, and the best way to proceed ideally is numerically. What becomes clear is that if only a few, well spaced sub-bands are occupied, then thermal broadening should be retained in the discussion if $k_B T$ is comparable to the sub-band separation. Unfortunately, without an explicit expression for τ_ϕ in a multi sub-band wire, it is difficult to be more quantitative than this on this point. In the following discussion the values quoted for L_{tr} / L_ϕ are therefore strictly valid only in the limit $T \rightarrow 0$ where thermal broadening is unimportant.

To illustrate the above ideas we have carried out simple calculations for three different experimental situations (the relevant parameters are either quoted by the authors or estimated from their findings). The appropriate value of the disorder is estimated from the elastic scattering times. The first device (Wheeler et al 1982) is a pinched silicon MOSFET; $W \sim 0.4\,\mu m$, $D \sim 1.4\times10^{-2} m^2 s^{-1}$, $l \sim 0.1\,\mu m$, $\tau \sim 7\times10^{-13} s$ and so the number of occupied sub-bands is of the order of 40. For this system $L_{tr}/L_\phi \sim 0.91$. They find that $L_\phi > 0.5\mu m$ for $T < 1K$, so $L_{tr} > W$ and the system will be one-dimensional for $T < 1K$. In the second system (Licini et al 1985), fine lithium wires were fabricated with widths ranging from $0.03\,\mu m$ to $25\,\mu m$. For one particular wire they

studied, $W \sim 0.074\,\mu m$, $D \sim 2.5\times10^{-2} m^2 s^{-1}$, $l \sim 0.05\,\mu m$ and $\tau \sim 10^{-13} s$. Since the Fermi energy of a metal is higher than that of a typically doped semiconductor, the number of occupied sub-bands is also higher, $\sim$ 100 for this wire. In this case $L_{tr}/L_\phi \sim 0.99$. The phase breaking length $L_\phi \sim 1.9T^{-1}\mu m$ so this system is easily $1D$ in the temperature range studied ($0.1<T<16K$). Both of these experiments involved many occupied sub-bands and so the usual criterion for $1D$ behaviour ($L_\phi \rangle W$) is sufficient. In the opposite limit (Thornton et al 1986), a very narrow, high mobility device was formed from a GaAs-AlGaAs heterojunction: $W \sim 0.045\,\mu m$, $D \sim 6.6\times10^{-2} m^2 s^{-1}$, $l \sim 0.7\mu m$, $\tau \sim 7\times10^{-12} s$ and $L_\phi \sim 0.2\,\mu m$ at $T = 0.5K$. For this system, the number of occupied sub-bands is only of the order of unity, and so we can anticipate $L_{tr} \ll L_\phi$ and $L_{tr} \ll l$. Notice however that the width of the channel is less than the mean free path, and so this is the length scale which controls the dimensionality in this case. The above calculations should not be taken too literally as many of the parameters used are only approximate, and one would have to have detailed information before taking them further. They do provide a useful guide to the type of behaviour we might expect however.

To sum up; we have used a simple model to illustrate possible sub-band effects in the weak localisation term. The one-dimensional nature of many of the integrals has meant that they can be evaluated exactly (using contour integration) which is an advantage. A very simple form arises for ΔG_{WL}, which is essentially independent of details of the sub-band structure, the precise form of the impurity scattering and the nature of the one-electron states. We therefore expect that the form of our results will also be found in a more sophisticated calculation. Criteria for one dimensionality have been given (and compared to existing criteria), and the explicit form of the weak localisation term in the intermediate regime between $1D$ and $2D$ has been calculated. Our results lead us to the tentative prediction that if channel widths are less than the mean free path, then the localisation will always appear one-dimensional, irrespective of inelastic effects. This result is pertinent to perfect walls (i.e. specular reflection). If the walls are rough however, and $W \sim l$, diffuse scattering will become important and the above model breaks down. Work is under way to tackle this situation using the methods proposed by Dugaev and Khmel'nitskii (1984) in disordered films, where the Cooperon is calculated from Boltzmann rather than the diffusion type of

equation considered by Al'tshuler and Aronov (1985) for the dirty channel case.

Finally we must clarify the relationship between the phase destroying events implicit in our treatment of the elastic scattering (the states are characterised by complex energies) and the supposed inelastic events. It is well known that the single-particle Green's function after configuration averaging contains no information about localisation in the system (Lee and Ramakrishnan 1985). Localisation only becomes apparent through the study of two-particle functions. The phase incoherence manifest in the complex self-energies of the single- particle functions should then be viewed as an artefact of the type of approximation we have used (a sort of coherent potential approximation) which does not really produce a physically observable phase relaxation. Although these Green's functions describe states which are not localised, the conjecture is that they may be used to investigate localisation through a study of the current-current correlation function, where the real observable incoherence due to inelastic processes is introduced artificially. This hidden assumption is behind all such treatments of the weak localisation problem. In principle, justification could arise from a study of the exact eigenstates of the disordered system, which have not been configurationally averaged (and so contain information about localisation) and which do not have complex energies.

CHAPTER 6:

THERMAL TRANSPORT IN DISORDERED SYSTEMS.

6.1 Weak localisation in thermal transport.

In this chapter, we wish to discuss the effects of disorder upon thermal transport, especially localisation and fluctuation phenomena. Unfashionable for a long time, this subject has recently gained impetus with several new and interesting effects being proposed (Sivan and Imry 1986, Kaiser 1987, Strinati and Castellani 1987, Esposito et al 1987, Kearney and Butcher 1988, Serota et al 1988). We shall find for example that the thermopower has weak localisation corrections, an important result since a diagrammatic calculation by Ting et al (1982) concluded that there was no weak localisation term in $2D$. To begin with we shall perform the calculations diagrammatically to correct this result (which to the best of my knowledge has gone uncorrected) and to show the form that the corrections to the thermal transport coefficients take in the weakly disordered regime. Later we shall extend the Chester-Thellung result (section 3.5) to discuss thermal transport in general, disordered systems, and also investigate fluctuation phenomena which ultrasmall devices might be expected to exhibit.

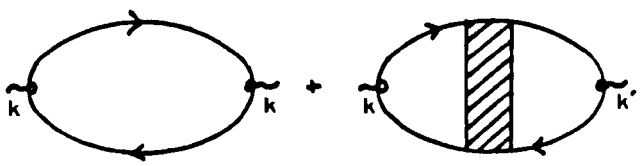
The Kubo formulae to be used in a calculation of the thermal transport coefficients were discussed in chapter 3. Each involves the heat current operator at least once, and therein lies the first problem, for this is a difficult quantity to calculate in general (Mahan 1981). For a free electron gas it has the form,

$$j_Q = \sum_k V_k \, \xi_k \, C_k^+ \, C_k \qquad (6.1)$$

where $\xi_k = \hbar^2 k^2/2m^* - \mu$. When extra (perturbation) terms are added in to the Hamiltonian however they also change the form of j_Q. If we wish to use the above form of the heat operator therefore we must be clear that the results are only accurate as the impurity concentration tends to zero. Later we shall discuss the situation for arbitrary disorder (in the absence of inelastic scattering), which will vindicate the decision to use equation (6.1) in a preliminary analysis.

The thermopower involves the correlation function L_{12} (equation 3.23). The dominant contributions after an $S-$ matrix expansion again come from the simple bubble and maximally

crossed diagrams,



where the unshaded circle represents the electric current vertex ($\propto -|e|$) and the shaded circle represents the heat current vertex ($\propto \xi_k$). The hatched region represents the amplitude for all the maximally crossed diagrams (the Cooperon). Both these diagrams are easily evaluated using the techniques of the previous two chapters. The contribution of the first may be written (in d dimensions),

$$L_{12}^{0} = -\frac{|e|\hbar^3}{m^{*2}} \frac{1}{L^d} \int \mathbf{k}^2 \left(\frac{\hbar^2 k^2}{2m^*} - \mu\right) d\mathbf{k} \int \frac{d\varepsilon}{2\pi} A(\mathbf{k}, \varepsilon)^2 \left[-\frac{\partial f}{\partial \varepsilon}\right] \tag{6.2}$$

where $A(\mathbf{k}, \varepsilon)$ is the spectral function and $f(\varepsilon)$ is the Fermi-function. In the twin limits $n_i \to 0$, $T \to 0$ this is easily evaluated to give,

$$L_{12}^{0} = -\frac{\pi^2}{3|e|} \frac{k_B{}^2 T^2}{\varepsilon_F} \sigma_B(\varepsilon_F) \tag{6.3}$$

where σ_B is the Boltzmann conductivity. To obtain the thermopower to this order we divide this result by $\sigma_B T$ and obtain, not surprisingly, the Mott result (equation 1.4). The next stage in the calculation is to evaluate the second diagram involving the Cooperon. In order to show that the result of Ting et al (1982) is incorrect, we shall explicitly consider $d = 2$. Most importantly when performing the calculation, we must remember that a finite temperature calculation involves an integral over the energy variable ε, and so the energy dependence of the diffusion constant and the Boltzmann conductivity must be retained. In the limit $n_i \to 0$, the spectral function becomes a δ– function and we derive (for an external field of frequency ω),

$$L_{12}^{1} = -\frac{2|e|}{\pi L^2} \int \sum_{Q} \frac{(\varepsilon - \mu) D(\varepsilon)}{D(\varepsilon) Q^2 - i\omega} \left[-\frac{\partial f}{\partial \varepsilon}\right] \tag{6.4}$$

At T=0 this term is zero, as of course all diagrams are for any thermal transport coefficient. At finite temperature and as $\omega \to 0$, we make the standard ansatz of replacing $i\omega$ by an effective cut-off $i\tau_\phi$ (the phase relaxation rate), which

is also a function of energy. If we assume that the phase coherence length L_ϕ has an energy dependence $\sim\varepsilon^y$, and that $k_B T \ll \varepsilon_F$ we find,

$$L_{12}^{\,1} = \frac{|e|}{6\hbar\varepsilon_F} k_B{}^2 T^2 (2y - 1) \tag{6.5}$$

Typically y will be rather small $\sim 2{,}3$ (Al'tshuler and Aronov 1985). The second diagram, since it has the same temperature dependence as the first in $2D$ and is of smaller magnitude (down by a factor of order $k_F l$) may thus be ignored. To evaluate the thermopower consistently to this order we must however divide by $(\sigma_B + \sigma_{WL})T$ whereupon,

$$S^{2D} = -\frac{\pi^2}{3|e|} \frac{k_B{}^2 T}{\varepsilon_F} [\, 1 - (k_F l)^{-1} \ln T/T_0] \tag{6.6}$$

with T_0 a reference temperature characteristic of the elastic scattering time. Hence there should be a weak localisation correction to the thermopower in $2D$ which disproves the result of Ting et al (1982). It is not totally clear why the discrepancy arises in their result since they present only an outline of the calculation, but it apparently has to do with their treating the energy dependences inconsistently. An experiment to check equation (6.6) would involve plotting S/T as a function of T to reveal the $\ln T$ dependence and would be well worth performing.

The extension to other dimensions is equally straightforward, and weak localisation corrections to the thermopower also exist in $3D$ (which supports the recent conjecture of Kaiser 1987) and $1D$. The behaviour in one dimension is interesting because at length scales greater than the mean free path the weak localisation term is comparable to the Boltzmann term, whereupon the thermopower should exhibit rather exotic behaviour. We shall return to this when we discuss fluctuations. The thermal conductivity can also be calculated diagrammatically in the same fashion. What one finds is that κ has same dependence on the conductivity, even when weak localisation corrections are taken into account, as it had to lowest order. In particular, there are no weak localisation corrections to the Wiedemann-Franz law. This is a consequence of a rather more general scaling theory of thermal transport from which all the above results can be derived and which we will now discuss.

6.2 The Chester-Thellung theorem revisited.

The Chester-Thellung theorem was discussed in some detail in section (3.5). Under the assumptions of elastic scattering and independent electrons, it enables a series of identities to be developed between the thermal transport coefficients which are true for arbitrary disorder. This is the point we now wish to examine a little more closely, with the aim of using the results to generate a plausible scaling theory of thermal transport. The first point to notice is that the theorem does not depend upon the nature of the eigenstates, and is thus true for localised as well as extended states. All the information about the disorder is contained in the conductivity. Secondly we notice a strong similarity between the results of perturbation theory just discussed and the exact expressions we have derived in section (3.5) for elastic scattering. There are two important conclusions to be drawn from this. Firstly the exact results in the absence of inelastic scattering and the perturbative results are equivalent in the limit of zero impurity concentration and so the diagrammatic analysis above is valid. Secondly, if the scattering is weakly inelastic we assume that the results still hold, since the main effect of such scattering is to destroy phase coherence rather than change the energy of the electrons. This is the assumption that lies behind the neglect of thermal broadening in the usual treatment of the weak localisation problem. We can then use the theorem to study weak localisation phenomena, and as confirmation of the validity of the approach we find the same answers as derived from perturbation theory. The weak localisation corrections are then given generally by inserting the correct expression for the electrical conductivity in equation (3.27a,b,c).

Study of the scaling theory of localisation highlighted the fact that the single variable needed to specify the behaviour was the conductance at a length scale L. The form of the above results strongly suggests that this is also essentially true for the thermal transport coefficients of non-interacting electrons. We shall therefore make the following plausible assumption. If we insert the relevant expression for the zero temperature conductance in the formulae rather than the conductivity, then this will lead to the relevant thermal transport parameters for a finite sized system, which are the thermal conductance and (still) the thermopower. This is borne out of the Landauer type arguments of Engquist and Anderson 1981 and Sivan and Imry 1986. The results will

be valid as long as inelastic scattering can be neglected, and in the forthcoming analysis this should always be kept in mind.

6.3 Scaling theories of thermal transport.

6.3.1 The thermal conductance.

When calculating the thermal conductance we want to keep the temperature low in order to minimise inelastic effects. Since no thermal transport is possible at T=0 however we have to be a little careful in handling this limit. At low temperature, we have seen that $\kappa \propto T$, and so the sensible parameter to consider therefore is κ/T as $T \rightarrow 0$. At sufficiently low temperature, the electrical conductance G will always be slowly varying on a scale of $k_B T$ enabling us to expand the $L_{\alpha\beta}$ (equation 3.27) and write

$$L_{11} = G(\varepsilon_F) \tag{6.7a}$$

$$L_{12} = L_{21} = -\frac{\pi^2}{3|e|}(k_B T)^2 G'(\varepsilon_F) \tag{6.7b}$$

$$L_{22} = \frac{\pi^2}{3|e|^2}(k_B T)^2 G(\varepsilon_F) \tag{6.7c}$$

where the prime denotes an energy derivative. The thermal conductance of a finite sample is simply $(L_{22}L_{11} - L_{12}L_{21})/TL_{11}$. Substituting in equations (6.7) enables us to write,

$$\kappa = \frac{\pi^2}{3|e|^2} k_B{}^2 T\, G(\varepsilon_F) + O\left[\frac{k_B T}{\gamma}\right]^2 \tag{6.8}$$

where γ is the energy scale over which G is approximately constant. In the limit $T \rightarrow 0$ therefore,

$$\lim_{T\to 0} \frac{\kappa}{GT} = \frac{\pi^2}{3|e|^2} k_B{}^2 \tag{6.9}$$

This is nothing other than the Wiedemann-Franz law, and this general proof of its validity was first given by Chester and Thellung (1961). What we are stressing is that κ/T is the natural scale variable which has exactly the same behaviour as the zero temperature conductance, and so all the results of the usual scaling theory (discussed at length in chapter 4) apply to this variable as well. The two most interesting consequences of this to test experimentally are that there are no weak localisation corrections to the Wiedemann-Franz law, and that in the vicinity of a Metal-Insulator transition, the thermal conductance has the same critical exponent as the conductivity,

$$\lim_{T\to 0}\frac{\kappa}{T}\sim(\varepsilon_F-\varepsilon_c)^{\nu} \tag{6.10}$$

This result was also hinted at by Strinati and Castelleni (1987) who used diagrammatic perturbation theory, though the proof given here is considerably simpler and more general in that no assumption about the disorder being weak is required. In order to test this experimentally it must be remembered that for a given ε_F arbitrarily close to the mobility edge we must have $k_BT \ll (\varepsilon_F - \varepsilon_C)$. It should certainly be possible to measure the thermal conductance near the mobility edge to confirm this result. One further aspect is that in an infinite $2D$ or $1D$ sample there is no thermal conductance in the absence of inelastic hopping, which reflects the fact that localised states which cannot carry current, cannot carry heat either. The equivalence of κ/T and G outlined above may of course be viewed in terms of this simple intuitive picture.

6.3.2 The thermopower.

The behaviour of the thermopower in disordered systems on the other hand turns out to be very different. The first attempt at a scaling theory of the thermopower was given by Sivan and Imry (1986), who made particularly elegant use of a generalised Landauer formula. All the results they derive can also be obtained from the Chester-Thellung theorem. The first point that arises we have already discussed; that there are weak localisation corrections to the thermopower. Once the system size exceeds the phase coherence length L_ϕ, both L_{12} and L_{11} have the same dependence on L, and so the thermopower becomes independent of system size (and in particular remains finite) in all dimensions as $L \to \infty$. This makes its behaviour very different from G and κ/T, both of which vanish for example in $2D$ and $1D$ as $L \to \infty$. The behaviour near the mobility edge is also predicted to be completely different. If $k_BT \ll \varepsilon_F - \varepsilon_c$ then we find,

$$S = \frac{-\pi^2}{3|e|}k_B\nu\frac{k_BT}{\varepsilon_F-\varepsilon_c} \tag{6.11}$$

As the mobility edge is neared the thermopower initially decreases. Eventually there will come a point however when $\varepsilon_F - \varepsilon_C$ is comparable to k_BT and we must revert to the full expressions (equation 3.27) for the $L_{\alpha\beta}$. When $k_BT \gg \mu - \varepsilon_C$ we find (c.f. Sivan and Imry 1986),

$$S = -\frac{k_B}{|e|}\left[A - B\,\frac{(\mu-\varepsilon_c)}{k_BT}\right] \tag{6.12}$$

where A and B are constants which depend upon the precise value of the critical exponent ν, and B is of order unity (though not exactly equal to one as stated by Sivan and Imry 1986). The result implies that if the chemical potential is varied so that it crosses a mobility edge, then S will vary smoothly across that mobility edge. In particular (and in sharp contrast to σ and κ), no structure is expected for $\mu \sim \varepsilon_c$. This would also be very interesting to test experimentally. One further point is that the value of the thermopower at the mobility edge is predicted to be a universal constant $-k_B A / |e|$ for all materials, which it should also be possible to investigate. The problem associated with the phonon-drag component can be reduced by working at low temperatures, and in very disordered samples where the phonon mean free path will be limited. In such systems a direct comparison may prove possible between the above results for the diffusive component of S and the experimentally measured values.

6.4 Fluctuation phenomena in thermal transport.

One important feature of the Chester-Thellung results is that no ensemble average has had to be taken in deriving them. The results are thus applicable to single samples, in particular, to small samples which are non-self-averaging. We have already discussed that the conductance of such systems exhibits sample specific fluctuations as a function of chemical potential at low temperatures (Lee et al 1987). The form of the above results leads us to the tempting conclusion that the analogue of these universal conductance fluctuations may also be present in the thermal transport coefficients. This has recently been investigated by Esposito et al 1987 (who invoked an extension of the correlation function ideas of Lee and Stone 1985) and Serota et al 1988 (who considered the problem in multi-lead devices). We shall pursue our use of the Chester-Thellung result applied to a single sample, and use only the facts that the zero temperature conductance fluctuates with universal amplitude $\sim e^2/h$ and mean peak spacing of E_c (see section 4.7).

To begin with we shall look at some analytic results. Let us suppose that the temperature is low; $k_B T \ll E_c$, whereupon the simple expansions in equation 6.7 are valid. The coefficient L_{22} behaves rather similarly to the electrical conductance in this case, and according to the hypothesis that κ/T and G are similar scaling variables, we anticipate the same kind of fluctuation effects in

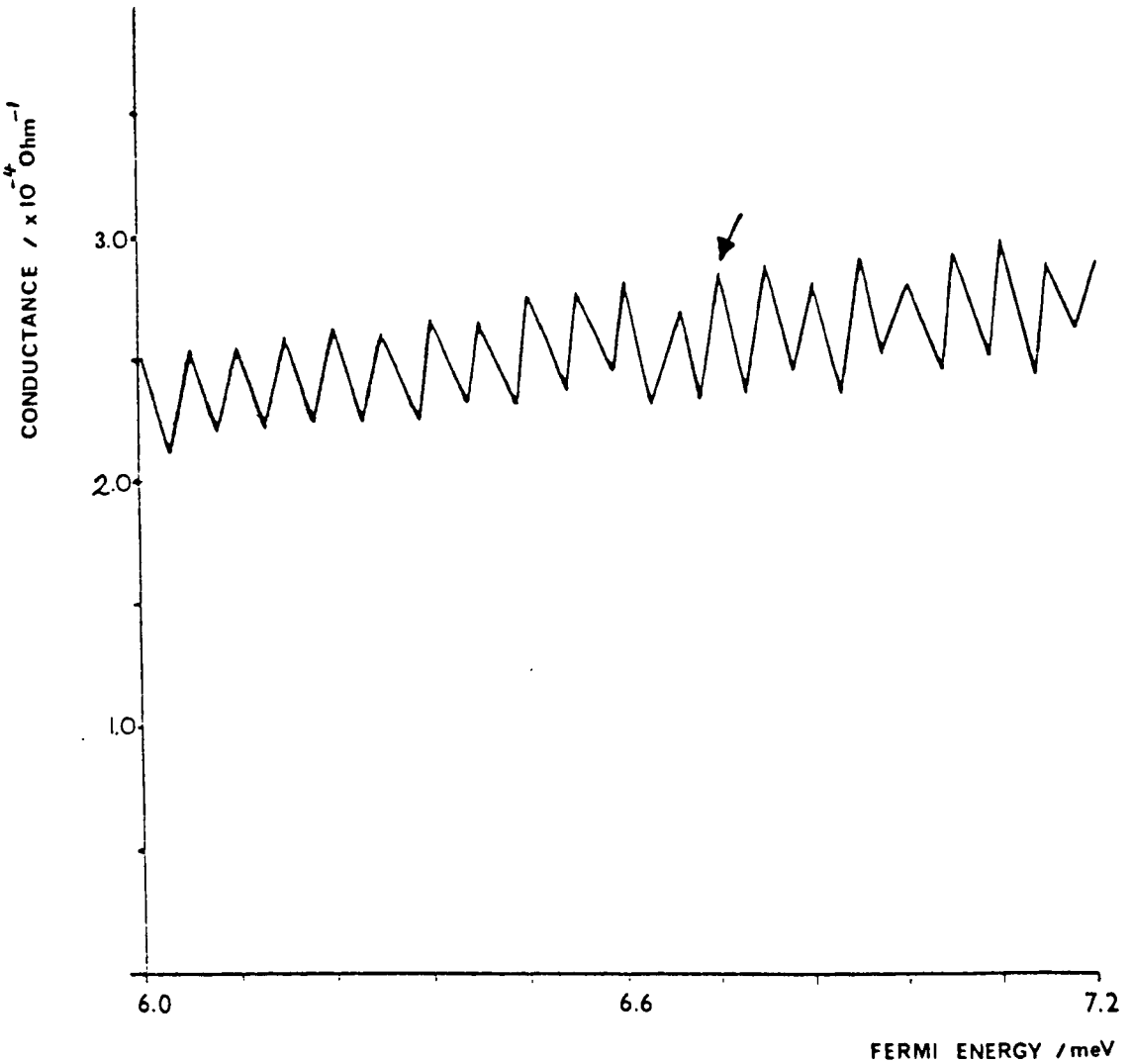


Figure 6.1: The zero temperature conductance as a function of Fermi energy. The fluctuations have amplitude $\sim e^2/h$ about a mean value of order $6e^2/h$. The arrowed peak is plotted as a function of temperature in figure 6.3.

the thermal conductance as the electrical conductance. The fluctuations in the coefficient L_{12} are more interesting however. An approximate expression for the derivative of the conductance is,

$$G'(\epsilon_F) = \pm a\frac{e^2}{h} \times \frac{1}{E_c} \qquad , a \sim 1 \tag{6.13}$$

and since it can have either sign (depending on the exact position of the Fermi energy) the thermopower will inevitably fluctuate wildly in sign and magnitude as the Fermi energy alters. The magnitude of the oscillations in L_{12} is simply,

$$\delta L_{12} = \frac{\pi^2}{3|e|}\frac{k_B e^2 a}{h}\frac{k_B T}{E_c} \tag{6.14}$$

The correlation energy $E_c = \hbar\pi^2 D/L_z^2$ and so depends upon the system size and the degree of disorder, and equation (6.14) holds as long as $k_B T \ll E_c$. For $k_B T > E_c$, thermal energy averaging begins to wash out the fluctuations in the thermal transport coefficients as we should expect. Esposito et al (1987) used this idea to say when $k_B T = E_c$ the magnitude of the L_{12} fluctuations is maximum and also universal. It is not clear however that this has any immediate consequence for the thermopower that one might measure in a device. In order to be more specific about this and also the thermal conductance fluctuations we therefore have to proceed numerically using the full results (equation 3.27).

The following results are pertinent to the situation where the sample size is less than L_ϕ, when the important effect is thermal averaging within the whole sample. We consider for the sake of argument a silicon strip of order $1\mu m$ long with a longitudinal diffusion coefficient of $D \sim 10^{-2} m^2 s^{-1}$. The mean conductance is chosen to be linear in Fermi energy with a magnitude of the order of $6e^2/h$, upon which we superimpose fluctuations $\sim e^2/h$ in a random fashion. The correlation energy E_c is $0.06meV$. The resulting zero temperature conductance is plotted as a function of Fermi energy in figure 6.1, and has a similar form to that measured in actual devices. As the temperature increases the fluctuations are washed out (figure 6.2) and when $k_B T \sim 0.5E_c$ they have essentially disappeared. In this case we are simply left with the mean conductance. The temperature dependence of one of the peaks in the conductance is shown in figure 6.3 and shows the importance of thermal broadening even at relatively low temperatures. Next we calculate the thermal conductance as a function of chemical potential (figure 6.4). When the temperature is low,

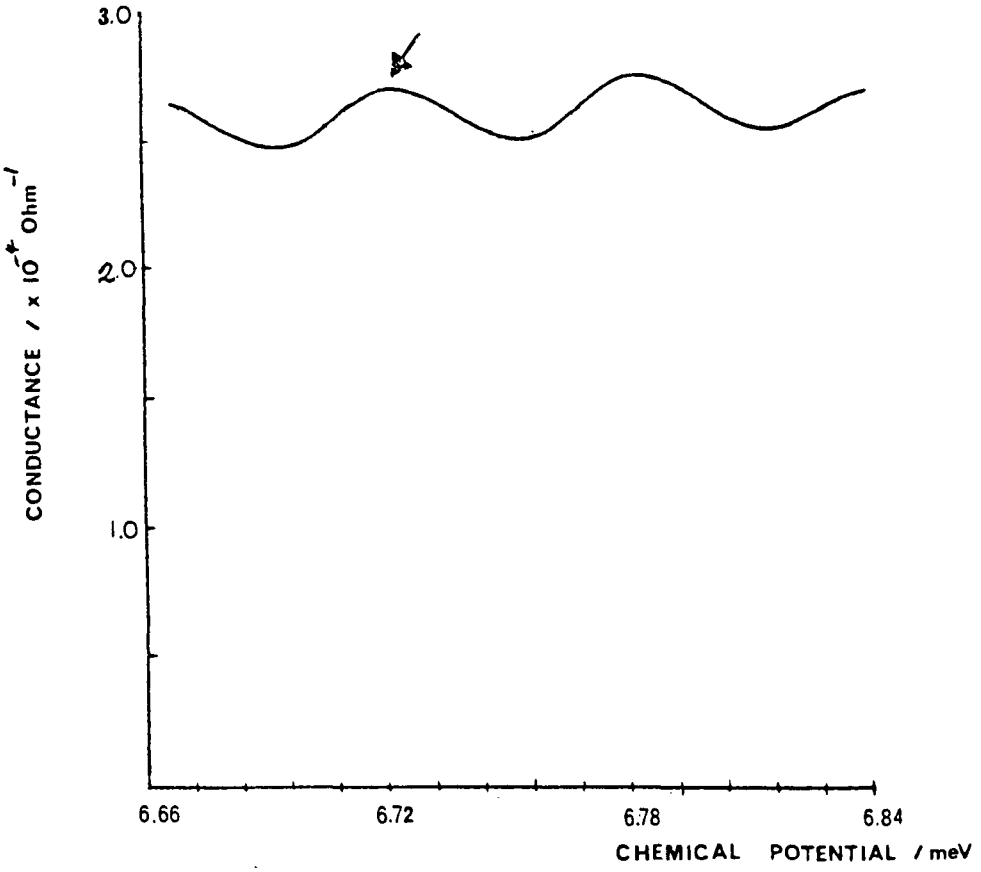


Figure 6.2: The conductance as a function of chemical potential at a temperature $k_B T = 0.1 E_c$, where E_c is the zero temperature correlation energy.

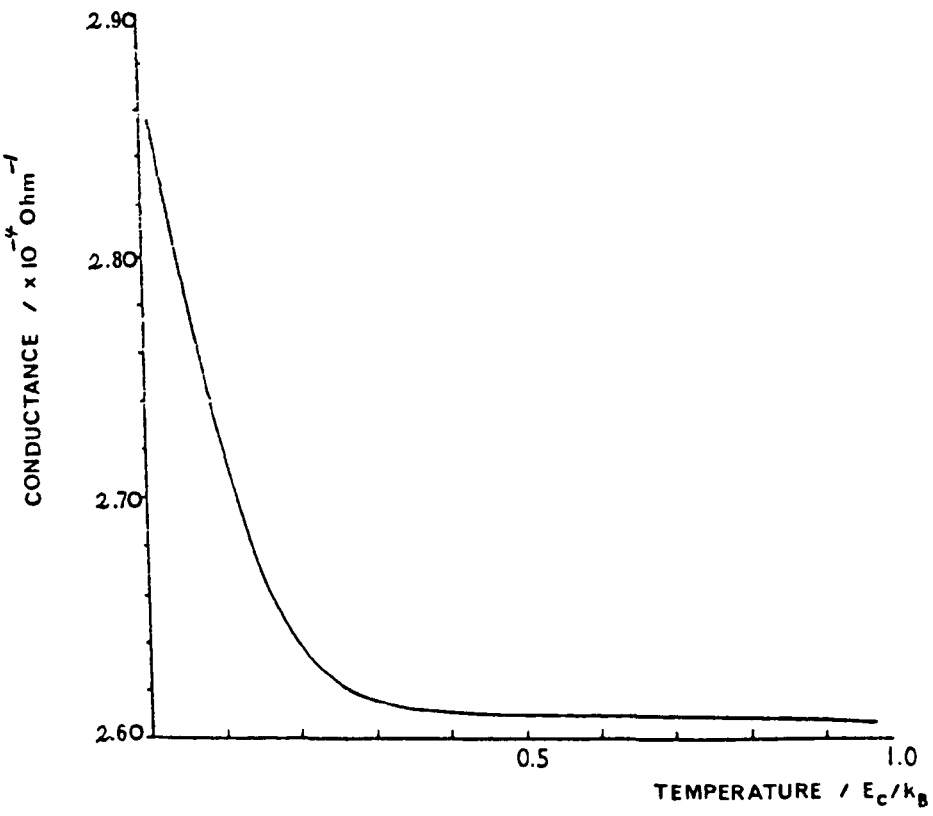


Figure 6.3: A plot of the temperature dependence of the indicated peak in the previous two figures.

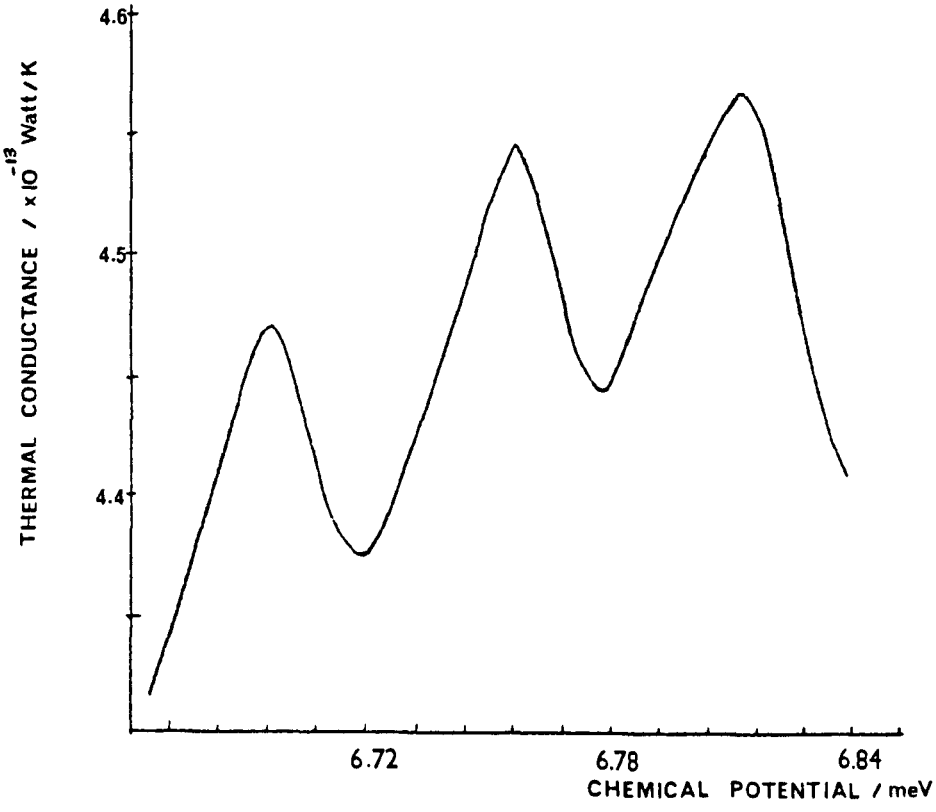


Figure 6.4: Thermal conductance as a function of chemical potential for $k_B T = 0.1E_c$.

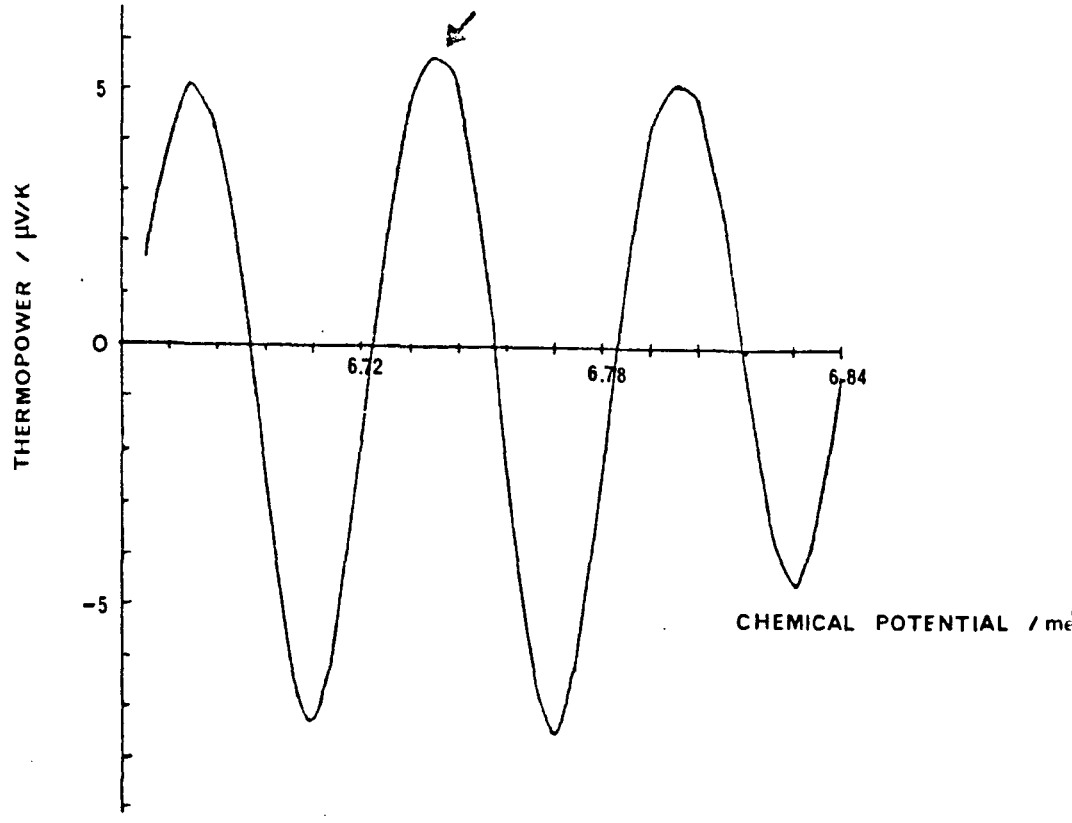


Figure 6.5: The thermopower as a function of chemical potential for $k_B T = 0.1E_c$.

fluctuations are evident, but these fluctuations are not universal in that they depend on temperature and also upon the exact value of the disorder (through E_c). As the temperature increases the peaks and troughs increase in a roughly linear manner but at different rates, so that when $k_B T \sim 0.5 E_c$, all evidence of the structure has vanished. The thermal conductance is then linear in temperature and has exactly the form we would expect. Above this temperature where the structure is coarse grained we recover the Wiedemann-Franz law, since $k_B T$ is now slowly varying on the relevant energy scale which is the chemical potential. It is important to realise though that at sufficiently low temperatures where $k_B T \ll E_c$, the Wiedemann-Franz law will also hold as well. This is nicely borne out by our results.

The fluctuations in the thermopower are much more interesting. At the lowest temperatures the expansion employed in deriving the Mott formula is valid, and as a result the sign of the thermopower at a given chemical potential will depend upon the sign of the energy derivative of the conductance. Since this is a rapidly fluctuating quantity we must expect the thermopower to change sign on a scale of $\sim E_c$, and this is shown in figure 6.5. The size of the fluctuations obviously depends upon the temperature in a subtle way, since initially they grow in T until thermal broadening becomes important, when they will begin to die away. As an illustration of this we have selected the peak indicated in figure 6.5 and followed its evolution as a function of temperature (figure 6.6). This positive peak attains its maximum value when $k_B T \sim 0.1 E_c$ after which it begins to die away rapidly. This suggests that the "universal" amplitude suggested by Esposito et al (1987) occurring for $k_B T = E_c$ does not have much relevance in an actual measurement of the thermopower. In the final graph (figure 6.7) we have plotted the thermopower as a function of chemical potential for $k_B T = 0.5 E_c$. This shows the demise of the fluctuation peaks which decrease more rapidly than the troughs rise, so the thermopower eventually becomes negative everywhere. At slightly higher temperatures the structure entirely disappears and the expected Mott result is recovered.

These results were applicable to samples where $L < L_\phi$. When the sample size exceeds this length, self-averaging of the fluctuations amongst the different sub-regions defined on a length scale L_ϕ begins to occur. For the conductance, the procedure was to assume that the different

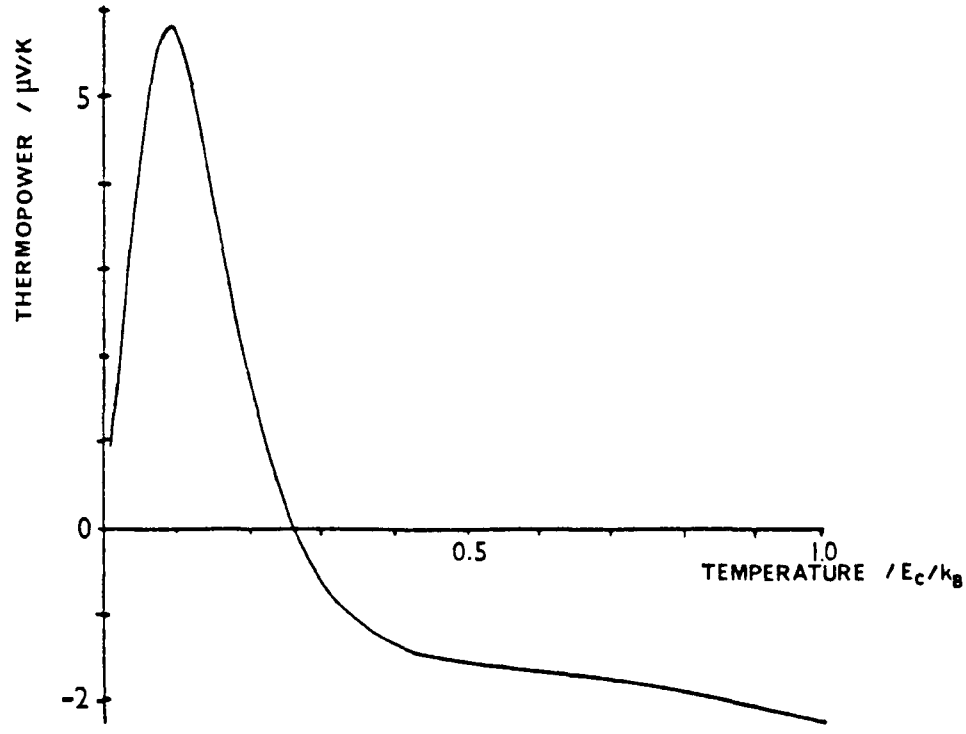


Figure 6.6: Evolution with temperature of the peak in the thermopower indicated in figure 6.5.

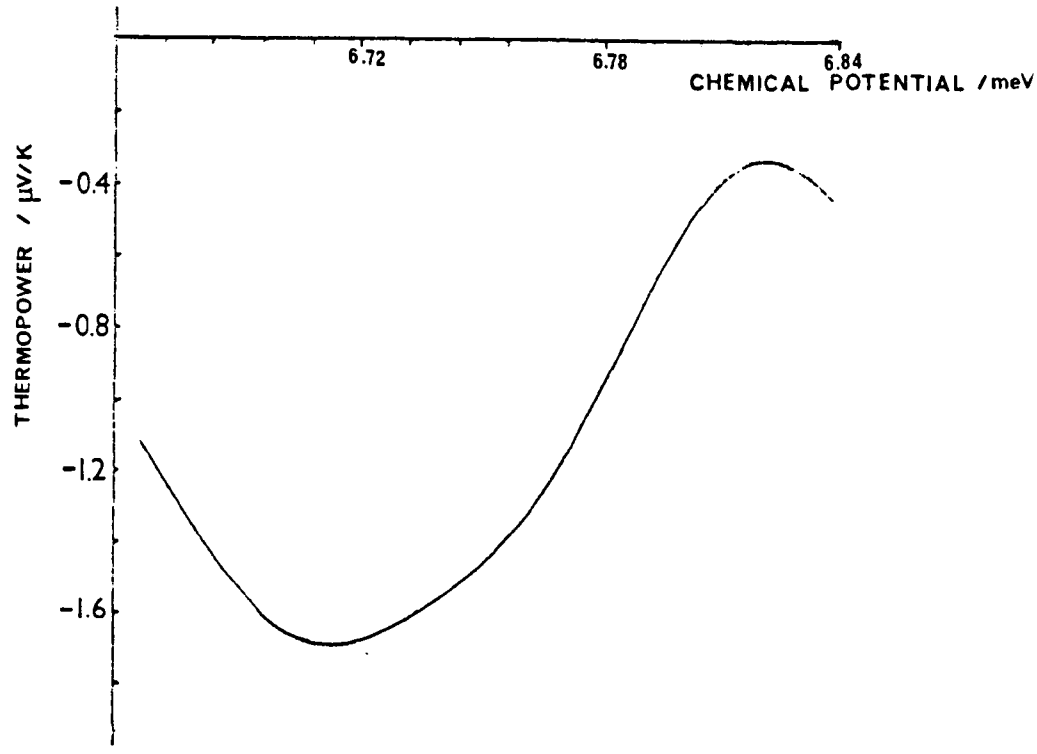


Figure 6.7: The thermopower as a function of chemical potential for $k_B T = 0.5E_c$.

random conductances of these sub-regins are essentially additive, whereupon the central limit theorem can be invoked (section 4.7) to show that the variance of the total conductance vanishes as $1 / L^d$. For the thermal transport coefficients however there is no simple procedure to repeat the argument, indeed deriving the classical combination laws for an ensemble of smaller samples is not trivial. As an example consider the total thermopower S of a parallel collection of smaller samples, each with thermopower S_i and conductance G_i. Under the assumption that conductance is additive one may show,

$$S = \frac{\sum_i S_i G_i}{\sum_i G_i} \tag{6.15}$$

There is no obvious use of the central limit theorem to discuss the variance of this quantity in terms of the distribution of its constituents. The same is also true of the thermal conductance. What one may show however is that both the thermopower and the thermal conductance do self-average in the limit $L \gg L_\phi$ although the exact scaling of the variance with L is not as yet clear. Work is under way to investigate this further. In order to maximise the fluctuations in the thermopower therefore the temperature should be made very low ($k_B T \ll E_c$), the sample small ($L < L_\phi$) and the mean conductance low as well. Our results indicate that it should certainly be possible to observe fluctuations in the thermopower of small wires and rings of semiconductor, and to a lesser extent in metallic samples. Although we have not proved it this will also be true of the magnetothermopower, which is one logical direction to develop the theory. The fluctuations should be sufficiently large to be observable despite the difficulties and uncertainties of measuring thermopower in such small devices.

CHAPTER 7: CONCLUSIONS AND FUTURE DEVELOPMENTS.

7.1 Summary and directions for improvement.

The previous six chapters represent an investigation into some aspects of electron transport in disordered systems, principally quasi-one-dimensional wires. By calculating the electrical conductance, thermal conductance and the thermopower in various regimes of interest, we have explored the possibilities of resolving QSE due to lateral quantisation, weak localisation corrections in multi sub-band systems as well as fluctuation phenomena in mesoscopic devices. What has emerged is that the thermopower is rather more sensitive to microscopic details of the system than the other two coefficients, and as such would constitute an interesting probe into the behaviour of devices now being realised. Many of the calculations we performed were of course rudimentary, although the formal theory necessary for more detailed work has also been given. It is therefore interesting to ask what are the logical improvements one could make to these calculations and how will they affect the results which we have presented in this work?

Initially we were concerned with the possibility of observing structure in the transport coefficients due to the presence of more than one occupied sub-band in the system. The conclusion was that lifetime broadening would have a significant effect, with only the thermopower proving likely to exhibit readily observable QSE. When considering the approximations used to derive these results it is natural to divide them into two classes; approximations made within the model and possibly important effects ignored by the model. In the former class, the major assumptions concerned the nature of the basis states (and with this the form of the confining potential), the neglect of the off-diagonal terms and the real part of the self-energy, and the δ–function form of the scattering potential. Of these, the first is the most trivial. It seems clear that an exact solution for the sub-band wavefunctions in the full confining potential will only alter details such as the position of the sub-band minima for example. Since QSE depend only upon there being structure in the density of states and not on the exact details of this structure, using the exact wavefunctions will probably only produce uninteresting quantitative changes to the transport coefficients. As such, the simple functions used would seem to be appropriate at this

level of calculation. The second and third approximations listed above are not quite so trivial however. It is certainly the case for example that if the sub-bands are close together, then the off-diagonal components and the real part of the self-energy will become more important. The question therefore is will this affect the transport coefficients in any major way? Perhaps the best way to look at this (short of a direct calculation), is to consider the model we proposed in chapter 5, which supposed periodic boundary conditions in the confined directions. In this model, the off-diagonal terms were found to be identically zero and so of course have no bearing upon the problem. The transport properties calculated within this model have a very similar behaviour to those calculated with more realistic (rigid) boundary conditions. It would be unreasonable to assume that the type of result we obtained was very sensitive to the choice of boundary conditions (Fisher and Lee 1981), and so we conclude that inclusion of the off-diagonal terms will not have that great an effect upon the problem. Inclusion of the real part of the self-energy $\mathrm{Re}\Sigma$ however would certainly produce changes if it turned out to be a significantly varying function of energy with a magnitude comparable to that of the imaginary part. $\mathrm{Re}\Sigma$ can be calculated either directly from the diagrams or from a suitably approximated imaginary component by invoking a Kramers-Kronig relation (Fetter and Walecka 1971). This is certainly the first logical step to take in improving the calculation. What is clear is that including $\mathrm{Re}\Sigma$ can only lead to a further broadening of the QSE in the transport coefficients. In the light of this, the results we have presented represent an upper bound to the amount of structure that can possibly be resolved in a experiment. The order to which the self-energy as a whole should be calculated is also an interesting point. Investigation of the structure of the other diagrams leads us to conclude that the approximation used (the self-consistent Born) is certainly the leading order contribution unless the impurity concentration becomes too high, when the matter of how to include the other diagrams is a universal problem, and not just restricted to the present problem. The important point is that the self-consistent Born approach is the minimal level of approximation that should be used; ignoring the self-consistency leads to singularities and divergences in the theory which cannot, in general, be reconciled.

The scattering mechanism we chose was the simplest possible that produces elastic scattering. For the Boltzmann term, the neglect of inelastic scattering is certainly going to be good at low temperatures. There are, however, many types of potential with which to model elastic impurity scattering. The δ– function form was chosen because, being zero range, it leads to a scattering vertex which is momentum independent, which in turn considerably simplifies the analysis. A potential of finite range does not have this property, but unless the potential is very long range this should not be too important. Quantitative changes can always be taken into account to a first degree by adjusting the strength of the δ– function in any case, so for the Boltzmann term the changes arising from using a more realistic scattering should only be minor. It would however be interesting to do a calculation with Yukawa-type potentials for example to see exactly how the finite range of the scattering enters the problem. If the range $R \ll k_F^{-1}$ then the effect should really be minimal.

Of the many things that our model neglects completely, we shall mention some of the potentially more interesting ones. Firstly we can imagine that most conducting quasi-one-dimensional systems are non-uniform, with walls that are rough on the atomic scale. This will lead to boundary scattering, which is important in other situations (see e.g. Dugaev and Kheml'nitskii 1984) and may be important here. A preliminary look at this has been taken by Van Houten et al (1988). Secondly, the effect of electron heating and non-linear transport may also be important, especially in ultra-small devices where maintaining a small electric field is not always easy. Finally, the inclusion of a magnetic field may help to resolve the sub-band structure by sharpening up the singularities in the density of states (Berggren et al 1986). To include such an effect within the theory we have outlined would be relatively straight forward.

In chapter 5, a calculation was performed for the weak localisation correction in a quantum wire, explicitly retaining the sub-band description of the basis states. Many of the ideas and criticisms mentioned above can also be levelled at this calculation, however the interference nature of the weak localisation phenomenon leads us to expect that the modifications resulting would only be of secondary importance. The exception to this concerns the nature of the impurity scattering which we shall return to in the next section. Inclusion of the real part of the self-energy for

example does not affect the essential structure of the theory, although its inclusion might lead to more concrete predictions concerning the transverse inelastic length (particularly its magnitude compared to the phase breaking length). One could also extend the theory to look at spin-orbit scattering in channels of arbitrary width, where the presence of another length scale in the problem could lead to interesting effects. The results of chapter 6 concerning thermal transport in disordered systems should be even more accurate, at least to the level of existing theories of the electrical conductance of disordered media. Since these are both quantitatively and qualitatively very good, we suggest that the ideas put forward concerning thermal transport should be directly verifiable by experiment as they stand. The simplest extension to the theory which leads to qualitatively different behaviour is to include a magnetic field. For the multi sub-band weak localisation model, this may permit analysis of the magnetoresistance of channels of varying width, provided that boundary scattering is unimportant. The validity of the Chester-Thellung theorem for magnetic fields in bulk solids (Smrcka and Streda 1977) suggests that all the existing theories of magnetic field effects in the weak localisation and fluctuation regimes will directly translate into the thermal transport coefficients. If this is the case, then the extensions are trivial and can be calculated immediately. The only remaining problem is to convince oneself that the Smrcka and Streda result is applicable in small geometry samples.

7.2 Future Developments.

In this final section, I would like to present some ideas I feel are promising areas for future work in disordered, low dimensional systems. Without doubt, the natural development experimentally is increasing miniaturisation, which will lead to more detailed analysis of fluctuations and related phenomena. Before considering these however, I would like to discuss some new aspects to a well established field, namely weak localisation. The traditional approach we have taken in this work relies on three fundamental assumptions concerning the impurities 1) they are static, 2) they have zero range in space and 3) they are uncorrelated in space. If any of these does not hold, in principle this may have an effect upon the weak localisation term. To see this, simply imagine some suitable (though absurd!) limits. If the impurity potential were infinite in range for

example (i.e. constant), then there would be no scattering and no weak localisation (it is possible however that this limit may be pathological). The same would apply if the impurities had sufficient spatial correlation i.e. if they defined a periodic lattice. Obviously if the impurities move, then this will lead to a dephasing effect which will also limit weak localisation. All these effects may then be thought of as producing some cut-off time in the problem, which, unlike the inelastic phase breaking time, will remain finite in the limit $T \rightarrow 0$, and must dominate the behaviour at sufficiently low temperatures.

There is certainly experimental evidence for such behaviour. In a careful experiment (Newson et al 1985), the phase breaking rate in InGaAs (a 3D sample) was extracted from low field magnetoresistance measurements and found to be of the form $A\,T^{3/2} + B$. The first term is due to electron-electron scattering, but the second does not have its origins in any inelastic mechanism. As an explanation, they offered the plausible idea that it might arise because of clustering amongst the impurities; a type of correlation. Can such effects as these be readily calculated? The question of dephasing due to the zero-point motion of the impurities has been addressed by Kumar et al (1987), who produced a simple estimate of the phase breaking time based on a Debye like, oscillatory model. Their result is of the right order of magnitude. Correlation effects however have only been been treated within a one-dimensional tight-binding model (Economou et al 1988). One way of including correlations within a more conventional diagrammatic treatment would be to introduce a model for the impurity distribution (with built in correlations) at the point where the ensemble average is taken. By assuming a completely random distribution we get of course momentum conservation within any one diagram. With an arbitrary distribution this is no longer the case, and so correlations will affect the divergence in the Cooperon and so limit weak localisation. A similar type of behaviour might possibly arise if we include finite range impurity scattering, where we must replace the structureless electron-impurity vertex U_0 by $U(\mathbf{q})$. The question is whether the delicate cancellation responsible for the divergence in the particle-particle channel then remains under this transformation. Work is currently proceeding to look at these interesting possibilities. Another aspect of the problem which has recently been studied (Fu 1988) concerns the interplay between localisation and fluctuations, which can lead (indirectly) to

a suppression of the localisation term.

The study of mesoscopic samples continues to produce surprising effects. We know for example that the conductance fluctuates from sample to sample as the impurity configuration is changed. This begs the question as to just how sensitive is the conductance of a given sample to the motion of a few impurities? This has been looked at by Feng et al 1986, Pelz and Clarke 1987, Feng et al 1987 who have shown that the motion of a single strong scatterer may lead to observable changes in the conductance of macroscopic samples. They suggested that this may even have implications for room temperature $1/f$ noise due to resistance fluctuations, and could provide a way of studying diffusion processes in solids with long timescales (of the order of minutes). Another important point which has been touched upon by the study of mesoscopics is the nature of a measurement. Most of the theoretical work has been done for a two probe geometry, with voltage and current measured in the same leads. Experimentally however, a four probe geometry is most commonly used. Normally in a classical system this would not produce any differences. In very small samples however where phase coherence can be maintained over the entire sample (including the leads), this is not the case, and the leads must be considered as part of the device rather than passive measuring probes. The conductivity of such a system is highly non-local, which can lead to voltage fluctuations as a function of magnetic field (independent of voltage probe separation up to a length L_ϕ) and asymmetries in the magnetoresistance. These interesting ideas have been treated theoretically, and also observed experimentally (see for example Buttiker 1986, 1987 and references therein). Effects such as these challenge the very nature of our concept of measurement, and will continue to be important in the future study of these devices.

Finally we should say a few words about ballistic transport. Devices can now be fabricated which are so small that essentially they are devoid of any impurities, and electrons can traverse the entire sample without suffering any collisions. A truly remarkable experimental demonstration of this has been given by Van Wees et al 1988 for point contacts defined in the 2DEG of a GaAs-AlGaAs heterojunction. They observe a conductance which is quantised in integer multiples of $e^2/\pi\hbar$, the integer referring to the number of transverse channels which are occupied. This

is exactly what is expected from an N-channel Landauer formula (with contacts included) if the transmission is perfect. This is very clear evidence that the transport through the point contact is almost completely ballistic. If indeed it does become possible to make small devices which are almost perfect, then the full range of exotic device possibilities such as the Aharonov-Bohm effect transistor (Bandyopadhyay et al 1986, Datta and Bandyopadhyay 1987) may become feasible. The physics of such systems also promises to be very interesting, since the quantum nature of the device will now be completely dominant. From a theoretical point of view, the challenge is obvious, since conventional theories based on diagrammatic methods will be completely inapplicable. In this respect, the Landauer formalism represents perhaps the most promising method for studying transport in the future.

APPENDIX: FORMAL IDENTITIES FOR A MULTI-SUB-BAND WIRE.

The general Matsubara function is,

$$G(\lambda,\nu,\tau) = -\langle T_\tau C_\nu(\tau) C_\lambda^\dagger(0)\rangle \tag{A1}$$

where $\langle \ldots. \rangle = Tr(\rho \ldots.)$, ρ is the density matrix and λ, ν both refer to the sub-band and momentum labels. If we consider the set of eigenstates of the operator $K = H - \mu N$,

$$K\,|n\rangle = E_n\,|n\rangle \tag{A2}$$

which are complete,

$$\sum_m |m\rangle\langle m| = 1 \tag{A3}$$

then we may write for $\tau > 0$,

$$G(\lambda,\nu,\tau) = -e^{\beta\Omega}\sum_n \langle n|\, e^{-\beta K}\, C_\nu(\tau) C_\lambda^\dagger(0)\,|n\rangle \tag{A4}$$

Inserting the identity (equation A3) between the creator and the annihilator gives,

$$G(\lambda,\nu,\tau) = -e^{\beta\Omega}\sum_{n,m} e^{-\beta E_n}\langle n|C_\nu|m\rangle\langle m|C_\lambda^\dagger|n\rangle e^{\tau(E_n - E_m)} \tag{A5}$$

The frequency transform is,

$$G(\lambda,\nu,i\omega_n) = \int_0^\beta G(\lambda,\nu,\tau)\, e^{i\tau\omega_n}\, d\tau \tag{A6}$$

$$\omega_n = (2n+1)\pi/\beta$$

and so,

$$G(\lambda,\nu,i\omega_n) = e^{\beta\Omega}\sum_{n,m}\frac{\langle n|C_\nu|m\rangle\langle m|C_\lambda^\dagger|n\rangle\,(e^{-\beta E_n} + e^{-\beta E_m})}{i\omega_n + E_n - E_m} \tag{A7}$$

The retarded function on the other hand is a real-time function,

$$G^R(\lambda,\nu,t-t') = -i\theta(t-t')\langle [C_\nu(t), C_\lambda^\dagger(t')]_+\rangle \tag{A8}$$

$$= -i\theta(t-t')\, e^{\beta\Omega}\sum_{n,m}\langle n|C_\nu|m\rangle\langle m|C_\lambda^\dagger|n\rangle\,(e^{-\beta E_n} + e^{-\beta E_m})\, e^{i(t-t')(E_n - E_m)} \tag{A9}$$

The Fourier transform of this function,

$$G^R(\lambda,\nu,\omega) = -i\int_0^\infty dt\, e^{i(\omega + i\delta)t}\, G^R(\lambda,\nu,t) \tag{A10}$$

where $\delta \to 0^+$, is then simply,

$$G^R(\lambda,\nu,\omega) = e^{\beta\Omega}\sum_{n,m}\frac{\langle n|C_\nu|m\rangle\langle m|C_\lambda^\dagger|n\rangle\,(e^{-\beta E_n} + e^{-\beta E_m})}{\omega + E_n - E_m + i\delta} \tag{A11}$$

Comparing equations A7 and A11 shows that the retarded function may be obtained quite gen-

erally from the Matsubara function through the analytic continuation $i\omega_n \rightarrow \omega + i\delta$. The advanced function may be obtained similarly by the continuation $i\omega_n \rightarrow \omega - i\delta$.

The spectral function $A(\lambda, \omega)$ is defined in terms of the diagonal components of the retarded Green's function:

$$A(\lambda, \omega) = -2 \operatorname{Im} G^R(\lambda, \lambda, \omega) \tag{A12}$$

and so,

$$A(\lambda, \omega) = 2\pi e^{\beta\Omega} \sum_{n,m} |\langle n | C_\lambda | m \rangle|^2 (e^{-\beta E_n} + e^{-\beta E_m}) \delta(\omega + E_n - E_m) \tag{A13}$$

The sum rule for $A(\lambda, \omega)$ is easily proved since,

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} A(\lambda, \omega) = e^{\beta\Omega} \sum_{n,m} |\langle n | C_\lambda | m \rangle|^2 (e^{-\beta E_n} + e^{-\beta E_m}) \tag{A14}$$

which may be written as,

$$= e^{\beta\Omega} \sum_n e^{-\beta E_n} \langle n | C_\lambda C_\lambda^\dagger + C_\lambda^\dagger C_\lambda | n \rangle = e^{\beta\Omega} \sum_n e^{-\beta E_n} = 1 \tag{A15}$$

The single particle density of states is obtained by considering the total number of particles in the system,

$$N = \sum_\lambda \langle C_\lambda^\dagger C_\lambda \rangle$$

$$N = e^{\beta\Omega} \sum_\lambda \sum_{n,m} e^{-\beta E_n} |\langle n | C_\lambda | m \rangle|^2 \tag{A16}$$

Now consider the integral,

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{1}{e^{\beta\omega} + 1} A(\lambda, \omega) \tag{A17}$$

Using the above expression for the spectral function we find this integral is,

$$= e^{\beta\Omega} \sum_{n,m} |\langle n | C_\lambda | m \rangle|^2 e^{-\beta E_n} \tag{A18}$$

Summing this over all λ then gives nothing more than the total number of particles in the system N (equation A16). Thus we have shown that,

$$N = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} f(\omega) \sum_\lambda A(\lambda, \omega) \tag{A19}$$

where $f(\omega)$ is the Fermi-Dirac function. It is therefore natural to associate the single-particle density of states with,

$$g(\omega) = -\frac{1}{\pi} \sum_\lambda \operatorname{Im} G^R(\lambda, \lambda, \omega) \tag{A20}$$

where we have used equation A12 for the spectral function.

REFERENCES.

Abrahams E, Anderson P.W, Licciardello D.C, and Ramakrishnan T.V 1979 Phys. Rev. Lett. 42 673

Abrahams E, Anderson P.W and Ramakrishnan T.V 1980 Phil. Mag. 42 827

Abrikosov A.A, Gor'kov L.P and Dzyaloshinskii I.Ye 1965 Quantum Field Theoretical Methods in Statistical Physics (Pergamon)

Abrikosov A.A and Ryzhkin I.A 1978 Adv. Phys. 27 147

Al'tshuler B.L, Khmel'nitskii D.E, Larkin A.I and Lee P.A 1980 Phys. Rev. B 22 5142

Al'tshuler B.L, Aronov A.G and Spivak B.Z 1981 J.E.T.P Lett. 33 101

Al'tshuler B.L, Aronov A.G and Khmel'nitskii D.E 1981 Solid State Commun. 39 619

Al'tshuler B.L and Aronov A.G 1981 J.E.T.P Lett. 33 499

Al'tshuler B.L, Aronov A.G and Khmel'nitskii D.E 1982 J. Phys. C 15 7367

Al'tshuler B.L, Aronov A.G and A.Yu Zyuzin 1984 Sov. Phys. J.E.T.P. 59 415

Al'tshuler B.L 1985 J.E.T.P. Lett. 41 648

Al'tshuler B.L and Aronov A.G 1985 in Electron-Electron Interactions in Disordered Systems ed. Efros and Pollak (North-Holland)

Al'tshuler B.L and Khmel'nitskii D.E 1986 J.E.T.P. Lett 42 359

Al'tshuler B.L 1987 Jpn. Jn. Appl. Phys. 26 1938

Anderson P.W 1958 Phys. Rev. 109 1492

Anderson P.W, Thouless D.J, Abrahams E and Fisher D.S 1980 Phys. Rev. B 22 3519

Anderson P.W 1981 Phys. Rev. B 23 4828

Ando T, Fowler A.B and Stern F 1982 Rev. Mod. Phys. 54 437

Aronov A.G and Yu.V Sharvin 1987 Rev. Mod. Phys. 59 755

Ashcroft N.W and Mermin N.D 1981 Solid State Physics (Holt-Saunders)

Azbel M.Ya 1980a J. Phys. C 13 L797

Azbel M.Ya 1980b Phys. Lett. 78A 410

Azbel M.Ya 1981 J. Phys. C 14 L225

Azbel M.Ya and Soven P. 1983 Phys. Rev. B 27 831

Bailyn M 1967 Phys. Rev. 157 480

Bandyopadhyay S, Datta S and Melloch M.R 1986 Superlatt. and Microstructures 2 539

Bate R.T 1988 Sci. Am. *March* P 78

Berezinskii V.L 1973 Sov. Phys. J.E.T.P. 38 620

Berggren K-F 1982a,b J. Phys. C 15 L45, L843

Berggren K-F and Newson D.J 1986 Semicond. Sci. Technol. 1 246

Berggren K-F, Thornton T.J, Newson D.J and Pepper M 1986 Phys. Rev. Lett. 57 1769

Bergmann G 1982 Phys. Rev. Lett. 48 1046

Bergmann G 1983 Phys. Rev. B 28 2914

Bergmann G 1984 Phys. Rep. 107 1

Bishop D.J, Dynes R.C, Lin B.J and Tsui D.C 1984 Phys. Rev. B 30 3539

Blatt F.J 1968 Physics of Electron Conduction in Solids (McGraw-Hill)

Board K 1985 Rep. Prog. Phys. 48 1595

Bonch-Bruevich V.L 1966 in Semiconductors and Semimetals, Vol. 1 (Academic)

Butcher P.N 1973 in Electrons in Crystalline solids ed. A Salam (Vienna: IAEA)

Butcher 1976 in Linear and Non-linear Electron Transport in Solids ed. Devrees and Van Doven (New York: Plenum)

Butcher P.N 1986 in Crystalline Semiconducting Materials and Devices (New York: Plenum)

Buttiker M, Imry Y, Landauer R and Pinhas S 1985 Phys. Rev. B 31 6207

Buttiker M 1986 Phys. Rev. Lett. 57 1761

Buttiker M 1987 Phys. Rev. B 35 4123

Cantrell D.G and Butcher P.N 1985a,b,c,d J. Phys. C 18 L587, 5111, 6639, 6627

Cantrell D.G and Butcher P.N. 1987a,b J. Phys. C 20 1985, 1993

Capasso F 1983a Surf. Sci. 132 529

Capasso F, Luryi S, Tsang W.T, Bethea C.G and Levine B.F 1983b Phys. Rev. Lett. 51 2318

Capasso F 1984 Surf. Sci. 142 513

Chester G.V and Thellung A 1961 Proc. Phys. Soc. 77 1005

Cutler M and Mott N.F 1969 Phys. Rev. 181 1336

Das Sarma S and Stern F 1985 Phys. Rev. B 32 8442

Das Sarma S and Xie X.C 1987 Phys. Rev. B 35 9875

Datta S and Bandyopadhyay S 1987 Phys. Rev. Lett. 58 717

Davidson J.S, Dan Dahlberg E, Valois A.J and Robinson G.Y 1986 Phys. Rev. B 33 8238

Davies R.A 1987 GEC Jn. Res. 5 65

Dolan G.J and Osheroff D.D 1979 Phys. Rev. Lett. 43 721

Dolan G.J, Licini J.C and Bishop D.J 1986 Phys. Rev. Lett 56 1493

Doniach S and Sondheimer E.H 1974 Green's functions for Solid State Physicists (W.A Benjamin Inc.)

Dugaev V.K and Khmel'nitskii D.E 1984 Sov. Phys. J.E.T.P. 59 1038

Economou E.N and Soukoulis C.M 1981 Phys. Rev. Lett. 46 618

Economou E.N and Soukoulis C.M 1981 Phys. Rev. Lett. 47 973

Economou E.N, Soukoulis C.M and Cohen M.H 1988 Phys. Rev. B 37 4399

Edwards S.F 1958 Phil. Mag. 3 1020

Efetov K.B 1985 J.E.T.P. Lett. 41 619

Efros A.L and Shklovskii B.I 1985 in Electron-Electron Interactions in Disordered Systems (North-Holland)

Engquist H-L and Anderson P.W 1981 Phys. Rev. B 24 1151

Eranen S and Sinkkonen J 1987 Phys. Rev. B 35 2222

Esposito F.P, Goodman B and Ma M 1987 Phys. Rev. B 36 4507

Feng S, Lee P.A and Stone A.D 1986 Phys. Rev. Lett. 56 1960

Feng S, Lee P.A and Stone A.D 1987 Phys. Rev. Lett. 59 1062

Fetter A.L and Walecka J.D 1971 Quantum Theory of Many-Particle Systems (McGraw-Hill)

Fisher D.S and Lee P.A 1981 Phys. Rev. B 23 6851

Fishman G 1987 Phys. Rev. B 36 7448

Fletcher R, Mann J.C, Ploog K and Weimann G 1986 Phys. Rev. B 33 7122

Fowler A.B, Hartstein A and Webb R.A 1982 Phys. Rev. Lett. 48 196

Fu Y 1988 Phys. Rev. Lett. 60 345

Fukuyama H 1985 in Electron-Electron Interactions in Disordered Systems (North-Holland)

Gallagher B.L, Gibbings C.J, Pepper M and Cantrell D.G 1987 Semicond. Sci. Technol. 2 L456

Gor'kov L.P, Larkin A.I and Khmel'nitskii D.E 1979 J.E.T.P. Lett. 30 229

Greenwood D.A 1958 Proc. Phys. Soc. 71 585

Grimes C.C and Adams G 1979 Phys. Rev. Lett. 42 795

Hartstein A, Webb R.A, Fowler A.B and Wainer J.J 1984 Surf. Sci. 142 1

Heiblum M and Eastman L.F 1987 Sci. Am. *Febuary* P 64

Herring C 1954 Phys. Rev 96 1163

Hikami S, Larkin A.I and Nagaoka Y 1980 Prog. Theor. Phys. Lett. 63 707

Jali V.M, Kubakaddi S.S and Bulimani B.G 1987 Phys. Stat. Sol. 142 K25

Jali V.M, Kubakaddi S.S and Bulimani B.G 1987 Phys. Stat. Sol. 144 739

Kaiser A.B 1987 Phys. Rev. B. 35 2480

Kaplan S.B and Hartstein A 1986 Phys. Rev. Lett. 56 2403

Kaveh M and Mott N.F 1981 J. Phys. C 14 L177

Kawabata A 1980 Solid State Commun. 34 431

Kawabata A 1981 J. Phys. Soc. Jpn. 50 2461

Kearney M.J and Butcher P.N 1986 J. Phys. C 19 5429

Kearney M.J and Butcher P.N 1987 J. Phys. C 20 47

Kearney M.J and Butcher P.N 1988 J. Phys. C 21 L265

Kearney M.J and Butcher P.N 1988 J. Phys. C in press.

Kelly M.J and Nicholas R 1985 Rep. Prog. Phys. 48 1699

Kelly M.J and Wiesbuch C 1986 The Physics and Fabrication of Microstructures and Devices (Springer-Verlag)

Kohn W and Luttinger J.M 1957 Phys. Rev 108 590

Kubakadi S.S and Bulimani B.G 1985 J. Appl. Phys. 58 3643

Kubo R 1956 Can. Jn. Phys. 34 1274

Kumar N, Baxter D.V, Richter R and Strom-Olsen J.O 1987 Phys. Rev. Lett. 59 1853

Landauer R 1970 Phil. Mag. 21 863

Langer J.S 1960 Phys. Rev. 120 714

Langer J.S and Neal T 1966 Phys. Rev. Lett 16 984

Langreth D.C and Abrahams E 1981 Phys. Rev. B 24 2978

Laughlin R.B 1983 Phys. Rev. Lett. 50 1395

Laux S.E, Frank D.J and Stern F 1988 Surf. Sci. 196 101

Lee K.L, Ahmed H, Kelly M.J and Wybourne M.N 1984 Electronic Lett. 20 289

Lee P.A 1984 Phys. Rev. Lett. 53 2042

Lee P.A and Stone A.D 1985 Phys. Rev. Lett. 55 1622

Lee P.A and Ramakrishnan T.V 1985 Rev. Mod. Phys. 57 287

Lee P.A, Stone A.D and Fukuyama H 1987 Phys. Rev. B 35 1039

Lee P.A 1987 Jpn. Jn. Appl. Phys. 26 1934

Licciardello D.C and Thouless D.J 1975a Phys. Rev. Lett. 35 1475

Licciardello D.C and Thouless D.J 1975b J. Phys. C 8 4157

Licini J.C, Dolan G.J and Bishop D.J 1985 Phys. Rev. Lett. 54 1585

Lin J.J and Giordano N 1987 Phys. Rev. B 35 545

Mahan G.D 1981 Many Particle Physics (New-York: Plenum)

Masden J.T and Giordano N 1982 Phys. Rev. Lett. 49 819

McInnes J.A and Butcher P.N 1985 J. Phys. C 18 L 921

Milligan R.F, Rosenbaum T.F, Bhatt R.N and Thomas G.A 1985 in Electron-Electron Interactions in Disordered Systems (North-Holland)

Milsom P.K and Butcher P.N 1986 Semicond. Sci. Technol. 1 58

Milsom P.K 1987 Ph.D Thesis, University of Warwick

Mott N.F and Twose W.D 1961 Adv. Phys. 10 137

Mott N.F, Pepper M, Pollit S, Wallis R.H and Adkins C.J 1975 Proc. Roy. Soc. A345 169

Mott N.F and Davies E.A 1979 Electron Processes in Non-Crystalline Materials (Oxford: Clarendon)

Nag B 1980 Electron Transport in Compound Semiconductors (Springer-Verlag)

Nakajima S, Toyozawa Y and Abe R 1980 The Physics of Elementary Excitations (Springer-Verlag)

Neal T 1967 Phys. Rev. 169 508

Nelson D.F, Miller R.C and Kleinmann D.A 1987 Phys. Rev. B 35 7770

Newson D.J, McFadden C.M and Pepper M 1985 Phil. Mag. B 52 437

Newson D.J, Pepper M, Hall H.Y and Marsh J.H 1985 J. Phys. C 18 L1041

Newson D.J and Pepper M 1985 J. Phys. C 18 L1049

Nicholas R.J 1985 J. Phys. C 18 L695

Paalanen M.A, Rosenbaum T.F, Thomas G.A and Bhatt R.N 1982 Phys. Rev. Lett. 48 1284

Paalanen M.A, Rosenbaum T.F, Thomas G.A and Bhatt R.N 1983 Phys. Rev. Lett. 51 1896

Pelz T and Clark J 1987 Phys. Rev. Lett. 59 1061

Pepper M 1977 Proc. Roy. Soc. A353 225

Pollak M and Ortuno M 1985 in Electron-Electron Interactions in Disordered Systems (North-holland)

Ruf C, Obloh H, Junge B, Gmelin E and Ploog K 1988 Preprint

Santhanam P, Wind S and Prober D.E 1984 Phys. Rev. Lett. 53 1179

Santhanam P 1987 Phys. Rev. B. 35 8737

Serota R.A, Ma M and Goodman B 1988 Phys. Rev. B 37 (preprint)

Schmid A 1974 Z. Phys. 271 251

Sernelius B.E, Berggren K-F, Tomak M and McFadden C 1985 J. Phys. C 18 225

Sharvin D. Yu and Yu. V Sharvin 1981 J.E.T.P. Lett. 34 272

Shklovskii B.I and Efros A.L 1984 Electronic Properties of Doped Semiconductors (Springer-Verlag)

Siggia E.D and Kwok P.C 1970 Phys. Rev. B 2 1024

Sivan U and Imry Y 1986 Phys. Rev. B 33 551

Skocpol W.J, Jackel L.D, Hu E.L, Howard R.E and Fetter L.A 1982 Phys. Rev. Lett. 49 951

Skocpol W.J, Jackel L.D, Howard R.E, Craighead H.G, Fetter L.A, Mankiewich P.M, Grabbe P and Tennant D.M 1984 Surf. Sci. 142 14

Smith A.C, Janak J.F and Adler R.B 1967 Electronic Conduction in Solids (New-York: McGraw-hill)

Smith C, Ahmed H, Kelly M.J and Wybourne M.N 1985 Superlatt. and Microstructures 1 153

Smrcka L and Streda P 1977 J. Phys. C 10 2153

Stedman G.E and Kaiser A.B 1987 J. Phys. C 20 3943

Strinati G and Castellani C 1987 Phys. Rev. B 36 2270

Sugihara Ko, Hishiyama Y and Ono A 1986 Phys. Rev. B 34 428

Takeshima M 1986 Phys. Rev. B 33 7047

Ting C.S, Houghton A and Senna J.R 1982 Phys. Rev. B 25 1439

Thornton T.J, Pepper M, Ahmed H, Andrews D and Davies G.J 1986 Phys. Rev. Lett. 56 1198

Thouless D.J 1973 J. Phys. C 6 L49

Thouless D.J 1977 Phys. Rev. Lett 39 1167

Thouless D.J 1981 Phys. Rev. Lett 47 972

Trzcinski R, Gmelin E and Quessier H.J 1986 Phys. Rev. Lett. 56 1086

Umbach C.P, Washburn S, Laibowitz R. B and Webb R.A 1984 Phys. Rev. B 30 4048

Van Houten H, Beenakker C.W.J, Van Wees B.J and Mooij J. E 1988 Surf. Sci. 196 144

Van Wees B.J, Van Houten H, Beenakker C.W.J, Williamson J.G, Kouwenhoven L.P, Van der Marel D and Foxon C.T 1988 Phys. Rev. Lett. 60 848

Vilenkin A and Taylor P.L 1978 Phys. Rev. B 18 5280

Vollhardt D and Wolfle P 1980 Phys. Rev. B 22 4666

Warren A.C, Antoniadis D.A and Smith H.I 1986 Phys. Rev. Lett. 56 1858

Wasburn S, Schmid H, Kern D and Webb R.A 1987 Phys. Rev. Lett. 59 1791

Webb R.A, Washburn S, Benoit A.D, Umbach C.P and Laibowitz R.B 1987 Jpn. Jn. Appl. Phys. 26 1962

Wheeler R.G, Choi K.K, Goel A, Wisnieff R and Prober D.E 1982 Phys. Rev. Lett. 49 1674

Whittington G.P, Main P.C, Eaves L, Taylor R.P, Thoms S, Beaumont S.P, Wilkinson C.D.W, Stanley C.R and Frost J 1986 Superlatt. and Microstructures 2 381

Wolfle P and Bhatt R.N 1984 Phys. Rev. B 30 3542